

Db 121 AACCAATCCAGAGAGCAGGCGTAAACCTTGTGATATGTCAGAGACCAATTCCAC 180
 Oy 223 GTGAGAGCCATCCAGGCGAATATCATATATTCATATGGAGCCTTTGTCGGAGCTGGG 282
 Db 181 GTCAGAGCATCCAGGCGAATATCATATATTCATATGGAGCCTTTGTCGGAGCTGGG 240
 Oy 283 CAGCTCCAGAGAGAGAGATCTCTGAAACCTGAGAGGAGGAGCTCTACCACTCA 342
 Db 241 CAGCTCCAGAGAGAGAGATCTCTGAAACCTGAGAGGAGGAGCTCTACCACTCA 300
 Oy 343 GAGAAAGAGAGAGAGAGATCTCTGAAATTTCCAGAGCCTGTTGCAACTGCT 402
 Db 301 GAGAAAGAGAGAGAGATCTCTGAAATTTCCAGAGCCTGTTGCAACTGCT 360
 Oy 403 GAGATCCAAATGTTAAAGTGAAGTGAATTTGTCCTCCAAAGTGAAGTGAAG 462
 Db 361 GAGATCCAAATGTTAAAGTGAAGTGAATTTGTCCTCCAAAGTGAAGTGAAG 420
 Oy 463 GAGACAAAGTTCATTTGATGCTTGAATTCAGAGATCATTTGCTGACGCTTTT 522
 Db 421 GAGACAAAGTTCATTTGATGCTTGAATTCAGAGATCATTTGCTGACGCTTTT 480
 Oy 523 TCCAG 546
 Db 481 TCCAG 504

RESULT 6

B0176246/c

LOCUS

DEFINITION

UI-M-DJ2-bwg-d-13-0-UI.s1 NIH_BMAP_DJ2 Mus musculus cDNA clone

ACCESSION

B0176246

VERSION

B0176246.1

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

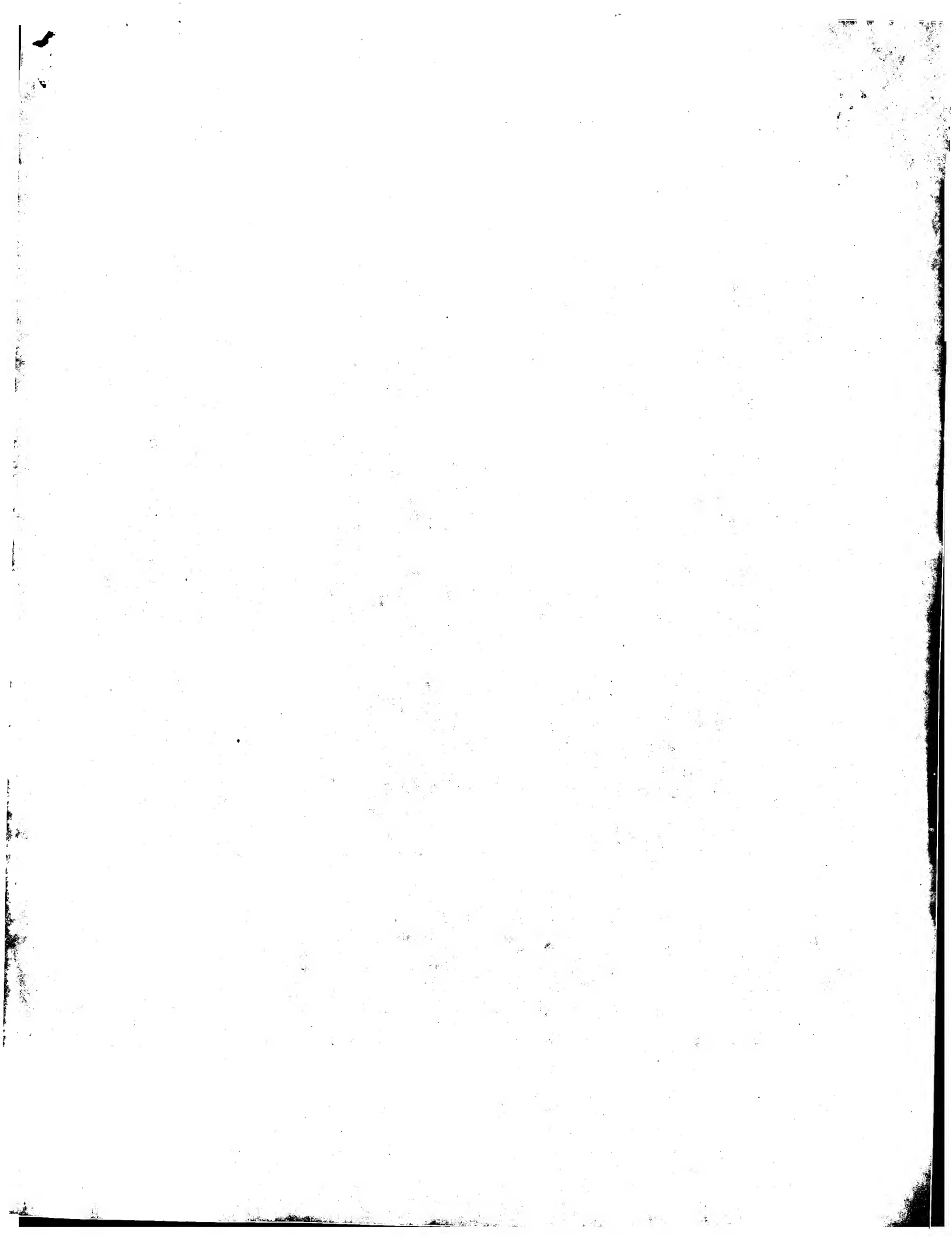
MEDLINE

COMMENT

FEATURES

source

1. 657
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-DJ2-bwg-d-13-0-UI"
 /clone_lib="NIH_BMAP_DJ2"
 /tissue_type="subcortical organ and postrema"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: brain; Vector: pT73-Pac (Pharmacia) with a
 657 bp mRNA linear EST 30-APR-2002
 UI-M-DJ2-bwg-d-13-0-UI.s1 NIH_BMAP_DJ2 Mus musculus cDNA clone
 B0176246
 B0176246.1 GI:20351738
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 657)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6(9):791-806 (1996)
 97044477
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 Tissue Procurement: Dr. Robin Davison
 CDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-40, >AT rich#low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLY-A=yes
 Location/Qualifiers
 1. 657
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-DJ2-bwg-d-13-0-UI"
 /clone_lib="NIH_BMAP_DJ2"
 /tissue_type="subcortical organ and postrema"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: brain; Vector: pT73-Pac (Pharmacia) with a



TAG_TISSUE=Subfornical organ and postrema
TAG_SEQ=GCTACATGAT"

Query Match	63.1%;	Score 491;	DB 14;	Length 657;
Beat [local] Stim[arty]	100.0%;	Prod No 150-113;		

288 TCCAGAGGAAAGAGTACTCTGAAACTGAGAGGGAGCTCTACCACTCAGAGA 347

Db 597 AAAGAGCCAA TTCCTGGATGAGAAATTTCCAGGACCTGTGTCACCTTGCTGAGAT 538

Db 537 CCAAATGTTAAAGTGAAGTGAATTTGTCCCAAGGTGAACAGTAGTCCGAAGACA 478

db 477 CAAAGTTCACATTTGGATGCTTAGAATCAGAGATGCATTTTCGTTCACGTCGTTTTCCAA 418

588 TGATCACAACTCTCTGAATGCCCAAGACTTACGCAAAATATCTGTTGTACATTAT 641

648 ATTTCTTCCTTTACTGTGTCATTCTCACTTAGCTACATTTTGGCACCTGTAGA 707

DY 708 GCAATCAGCACCGAATTACAACTGGGAAGTGTGTTTGAGGAGAGATGATTTT

Pb 237 GCAATCAGCACCGAATTACAACTGGGAAGTGTGTTTGAGGAGAGATGATTTT

Db 177 TATGAAGGGG 167

AK009857	LOCUS	AK009857	540 bp.	mrna	linear	HTC 19-JAN

ACCESSION
AK009857
VERSION
AK009857.1 GI:12844912
Insert sequence.

ORGANISM
Mus musculus
clone_11b:RIKEN full-length enriched mouse cDNA library
clone:2310046N05.

Phylum: Eukaryota; Chordata; Vertebrata; Euteleostei

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ. ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
type: nucleic acid

Query Match	5.6%;	Score 43.4;	DB 1;	Length 7218;
Best Local Similarity	3.6%;	Pred. No. 0.0036;		
Matches	11;	Conservative 173;	Mismatches 119;	Indels 0;
			Gaps	0;

Qy	256	ATGGAGACCTTTGCTCGGAGACTGGGACCTCCAGAGAAAGAGTACTCTGAA	315
Db	1451	ATGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	1392
Qy*	316	ACTGAGGAGGAGACTCTCAACACTAGAGAAAGAACCAATCTTGAATGAGAAA	375
Db	1391	RRR	1333
Qy	376	TTTCCAGGACCTGTGTCAACTGTGTGAGATCCAAAAGTTAAAGTGAAGTGAATT	435
Db	1331	RRR	1272
Qy	436	GTCGCCAAGGTGACAGTAGTGAAGGACAAAGTTCATGTGAGTGTGAAATC	495
Db	1271	RRR	1212
Qy	496	AGAGAGATGCTTTCGTTGACGTGTTTTCCAGGAGAAAGAAAACATGGGTGAATATA	555
Db	1211	RRR	1152
Qy	556	CAA 558	
Db	1151	RRR 1149	

RESUL 3
US-08-791-849A-15

Sequence 15, Application US/08791849A
Patent No. 5914449
GENERAL INFORMATION:
APPLICANT: Makoto MURASE et al.
TITLE OF INVENTION: Method for Increasing Storage
TITLE OF INVENTION: Lipid Content in Plant Seed
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/791,849A
FILING DATE: January 30, 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:

Qy 533 AAAAAACATGGCTGAAATTAACAACCTTCTGTAACCTTTATACATTTGTATGATGATGATC 592

Db 4135 AAAATGATGATTTTGTATACCCAGTTTTTGTATTTTGTATTTGACACTTTCTACAAAT 419

Qy 593 ACAACCTCTGATATGCCACAGACTCTAGCAAAAATTTCTGTGTTTGAACATTATATATTC 652

Db 4195 GTGAACTTTATTAATACAAAACCTTCAGGCTAAACATCCAAATTTTCTTTAATGCTTT 425

Qy 653 TTCCTTTTACTGTGCAATTTCTGACTTTAGTACATTTTTGGCACCTTTGAGA 707


```

? TITLE OF INVENTION: Aspartacylase Gene, Protein, and
? TITLE OF INVENTION: Methods of Screening for Mutations Associated with Canavan
? TITLE OF INVENTION: Disease
? NUMBER OF SEQUENCES: 68
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Millen, White, Zelano & Bryanigan, P.C.
? STREET: 2200 Clarendon Boulevard, Suite 1400
? CITY: Arlington
? STATE: Virginia
? COUNTRY: U.S.A.
? ZIP: 22201
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US94/07430
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/128,020
? FILING DATE: 29-SEP-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Hamlet-King, Diana
? REGISTRATION NUMBER: 33,302
? REFERENCE/DOCKET NUMBER: SHUTT 1PO
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 703-243-6333
? TELEFAX: 703-243-6410
? TELEX: 64191
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1435 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 159..1097
? PCT-US94-07430-1
?
? Query Match 4.3%; Score 33.8; DB 5; Length 1435;
? Best Local Similarity 55.6%; Pred. No. 1.3; 52; Indels 0; Gaps 0;
? Matches 65; Conservative 0; Mismatches
?
? Oy 552 TAAACAACCTCCGACACATTTATACATTTGATGATGATCACAACCTCCTGAATGCC 611
? DB 1233 TAACCAAAATTTCTAAATTAATTAATTAATCTTAAAGATATCATATTTAATGATGTGAG 1292
? Oy 612 AAGACTAGCAAAATAATATCTGTTGATACATTTAATTTCTCCCTTTACTGGGT 668
? Db 1293 CTATATCAAAAGAGTGTCTCTATTTCTATATATGTTATTATATACATGATCTTGAGT 1349
?
? RESULT 10
? PCT-US95-02251-2/c
? Sequence 2, Application PC/TUS9502251
? GENERAL INFORMATION:
? APPLICANT:
? TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
? TITLE OF INVENTION: CELLS
? NUMBER OF SEQUENCES: 18
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Arnold, White & Durkee
? STREET: P.O. Box 4433
? CITY: Houston
? STATE: Texas
? COUNTRY: United States of America
? ZIP: 77210
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patentin Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02251
FILING DATE: CONCURRENTLY HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC009P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3753 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3753
PCT-US95-02251-2

Query Match
Best Local Similarity 4.3%; Score 33.8; DB 5; Length 3753;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 56 CACTGGAATTGAGATCCCGCTCAGAGGACACCGGAGTTCTTATCTGTAAAGCGC 115
DB 2410 CACAGCGGCTCCGATCCCTTGACAGATGATAGCGGAGAGGACACTGGAAGAGC 2351
QY 116 TTTTGTGTTTGGACCTGCGCGCTGGGACTGTCTCAGGCAGTAACCAATCA 172
DB 2350 CTGGTGTGTTGTCGACAGATGCATCTTGGCACTTTCCACGCTTCACAGTATCA 2294

RESULT 11

US-08-199-780-2/c
Sequence 2, Application US/08199780
Patent No. 5763416

GENERAL INFORMATION:

APPLICANT: Bonadio, Jeffrey
APPLICANT: Goldstein, Steven A.
TITLE OF INVENTION: Gene Transfer Into Bone Cells
TITLE OF INVENTION: And tissues
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Parker, David
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4314 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 157..3912
US-08-199-780-2

Query Match
Best Local Similarity 4.3%; Score 33.8; DB 1; Length 4314;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 56 CACTGGAATTGAGATCCCGCTCAGAGGACACCGGAGTTCTTATCTGTAAAGCGC 115
DB 2569 CACAGCGGCTCCGATCCCTTGACAGATGATAGCGGAGAGGACACTGGAAGAGC 2510
QY 116 TTTTGTGTTTGGACCTGCGCGCTGGGACTGTCTCAGGCAGTAACCAATCA 172
DB 2509 CTGGTGTGTTGTCGACAGATGCATCTTGGCACTTTCCACGCTTCACAGTATCA 2453

RESULT 12

US-08-316-650-2/c
Sequence 2, Application US/08316650
Patent No. 5942496

GENERAL INFORMATION:

APPLICANT: Bonadio, Jeffrey
APPLICANT: Roessler, Blake J.
APPLICANT: Goldstein, Steven A.
APPLICANT: Lin, Mushan
TITLE OF INVENTION: METHODS AND COMPOSITIONS
TITLE OF INVENTION: FOR STIMULATING BONE CELLS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 30-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 4314 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 157..3912
US-08-316-650-2

Query Match 4.3%; Score 33.8; DB 2; Length 4314;
Best Local Similarity 55.6%; Pred. No. 2.3;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 56 CACGTGGAATTGAGTCCCGCTCAGAGACACCGGAGTTCTTCTATCTCTTAAGCC 115
DB 2569 CACAGCGGCTCGATCCCTTGACAGATATATGCGGAGAGCAGTCACTGAAAGAGC 2510
QY 116 TTTTGTGTTTGTGACCTGGCGCGCTGGAGCTGTCTCAGCAGTAACCAATCCA 172
DB 2509 CTGGTGTGTTCTGCAATGCATCTTGGCACACTTCCGAGCTCAGATCATCCA 2453

RESULT 13

US-08-928-419-5
Sequence 5, Application US/08928419
Patent No. 5859183
GENERAL INFORMATION:
APPLICANT: NAKAJIMA, YUKI
APPLICANT: YAMAMOTO, TOSHIYA
APPLICANT: OEDA, KENJI
TITLE OF INVENTION: CYTOPLASMIC MALE STERILITY DNA FACTOR
TITLE OF INVENTION: AND UTILIZATION THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,419
FILING DATE: 12-SEP-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, RAYMOND C.
REGISTRATION NUMBER: 21,066
REFERENCE/DOCKET NUMBER: 2185-0206P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 1..480
US-08-928-419-5

Query Match 4.3%; Score 33.6; DB 2; Length 480;
Best Local Similarity 52.1%; Pred. No. 0.88; Indels 0; Gaps 0;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 296 GGAAGAGAGTACTCTGAACTGAGAGGAGCTCTTACCACTTCAGAGAAAGAGC 355
DB 106 GGGATCAGCAGAAATTTAAACTGGGAAACCAATCTTTCACACAGAGAAACAATC 165
QY 356 CAATTCCTGAAATGAAGAAATTTCCAGAGCTGTGTCACTTGTCTGATCCAAATG 415
DB 166 CGAGCAGAGAGACCCCAACGATTGGAAAGATATCTTGAGAAAAGTTTAGACCCGATGA 225
QY 416 TTAAGTGAACGAAATTTGTC 439
DB 226 TCCTATATGATCTCCAGTTTATTC 249

RESULT 14

US-09-291-283-5
Sequence 5, Application US/09291283
Patent No. 6297012
GENERAL INFORMATION:
APPLICANT: NAKAJIMA, YUKI
APPLICANT: YAMAMOTO, TOSHIYA
APPLICANT: OEDA, KENJI
TITLE OF INVENTION: CYTOPLASMIC MALE STERILITY DNA FACTOR
TITLE OF INVENTION: AND UTILIZATION THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/291,283
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/928,419
FILING DATE: 12-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, RAYMOND C.
REGISTRATION NUMBER: 21,066
REFERENCE/DOCKET NUMBER: 2185-0206P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 1..480
US-09-291-283-5

Query Match 4.3%; Score 33.6; DB 4; Length 480;
Best Local Similarity 52.1%; Pred. No. 0.88; Indels 69; Gaps 0;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 296 GGAAGAGAGTACTCTGAACTGAGAGGAGCTCTTACCACTTCAGAGAAAGAGC 355
DB 106 GGGATCAGCAGAAATTTAAACTGGGAAACCAATCTTTCACACAGAGAAACAATC 165
QY 356 CAATTCCTGAAATGAAGAAATTTCCAGAGCTGTGTCACTTGTCTGATCCAAATG 415
DB 166 CGAGCAGAGAGACCCCAACGATTGGAAAGATATCTTGAGAAAAGTTTAGACCCGATGA 225

OY 416 TTAAGTGAAGTGAATTGTCC 439
Db 226 TCTATATGACTCCAGTTTATTC 249

RESULT 15

US-08-928-419-3
; Sequence 3, Application US/08928419
; Patent No. 5959183
; GENERAL INFORMATION:
; APPLICANT: NAKAJIMA, YUKI
; APPLICANT: YAMAMOTO, TOSHIBA
; APPLICANT: OEDA, KENJI
; TITLE OF INVENTION: CYTOPLASMIC MALE STERILITY DNA FACTOR
; TITLE OF INVENTION: AND UTILIZATION THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,419
; FILING DATE: 12-SEP-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, RAYMOND C.
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 2185-0206P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..651
; US-08-928-419-3

Query Match 4.3%; Score 33.6; DB 2; Length 651;
Best Local Similarity 52.1%; Pred. No. 1;

Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 296 GGAAGAGAGTACTCTGAACTGAGAGGAGCTCTACACCTCAGAGGAAAGAAC 355
Db 106 GGATTCAGCAGAAATCTTAAACTGGGACCACTTCTTTCACACGAGAGACACATC 165
OY 356 CAATTCTTGAATGAAGAAATTTCCAGACCTGTTGTCACTTGTCTGAGATCCAAATG 415
Db 166 CGAGCAAGGAGCCCAACAGTTTGAGAGATATCTTGAGAAAGGTTTACACCGGTGTA 225
OY 416 TTAAGTGAAGTGAATTGTCC 439
Db 226 TCTATATGACTCCAGTTTATTC 249

Search completed: April 15, 2003, 22:41:51
Job time : 75.0337 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 15, 2003, 16:52:31 ; Search time 1613.62 Seconds
(without alignments)
7608.593 Million cell updates/sec

Title: US-09-647-019-1

Perfect score: 1 ggcctcagcagcagcagagag.....ctgtatcttcatcagagggg 778

Sequence: 1 IDENTITY NUC

Scoring table: Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: em_ectba:*
2: em_ecthm:*
3: em_ectin:*
4: em_ectmu:*
5: em_ectov:*
6: em_ectpl:*
7: em_ectro:*
8: em_ectc:*
9: gb_ect1:*
10: gb_ect2:*
11: gb_ect3:*
12: gb_ect4:*
13: gb_ect5:*
14: gb_ect6:*
15: em_ectfun:*
16: em_ectom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_iny:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	776.4	99.8	934	11 AK003105	Mus muscu
2	770	99.0	934	11 AK010172	AK010172 Mus muscu
3	710	91.3	890	9 AV088480	AV088480 AV088480
4	608.4	78.2	613	14 B0554133	B0554133 B4026C08-
5	504	64.8	504	9 AA033164	AA033164 m137c11.r
6	491	63.1	657	14 B0176246	B0176246 UI-M-DJ2-

7	467.8	60.1	540	11 AK009857	AK009857 Mus muscu
8	456.4	58.7	617	12 BG794218	BG794218 UTSW_SMIG
9	453.4	58.3	482	9 AA34782	AA34782 ve23c01.r
10	448	57.6	486	9 A1035961	A1035961 ub50b12.r
11	430	55.3	578	14 B0554132	B0554132 B4026C08-
12	425.4	54.7	490	9 AA060214	AA060214 m165h06.r
13	425	54.6	436	14 W36988	W36988 mb5b11.r1
14	425	54.6	436	13 BM123288	BM123288 L0522D07-
15	401.6	51.6	448	14 W13738	W13738 mb32a12.r1
16	399.6	51.4	448	14 W18392	W18392 mb8b01.r1
17	389	50.0	369	9 A1098485	A1098485 uc06c10.r
18	386.6	49.7	468	14 W29828	W29828 mc07b11.r1
19	377.2	48.5	542	12 BF395174	BF395174 UI-R-CM0-
20	375	48.2	465	14 W18646	W18646 mb98a05.r1
21	367	47.2	478	9 A1613607	A1613607 vg04e07.y
22	366.8	47.1	437	9 A1613607	AA495581 vg04e07.y
23	365.4	47.0	402	10 AM743181	AM743181 up63f10.y
24	364	46.8	364	14 W97451	W97451 mf97c02.r1
25	362.6	46.6	416	9 A1153970	A1153970 ud49n02.r
26	362.6	46.6	630	12 BG791844	BG791844 UTSW_H1E3
27	361.8	46.5	521	9 AV005538	AV005538 AV005538
28	361.2	46.4	719	14 BM697544	BM697544 UI-E-DX0-
29	357	45.9	368	14 W29186	W29186 mc22f02.r1
30	352.4	45.3	459	9 AA800221	AA800221 E87189718
31	351	45.1	491	12 BF284886	BF284886 E87449487
32	350	45.0	507	12 BF395391	BF395391 UI-R-CM0-
33	349.4	44.9	510	10 AM918749	AM918749 B8750053
34	346.4	44.5	450	9 AA800829	AA800829 E87190326
35	324	41.6	756	12 BF791178	BF791178 602251278
36	320.4	41.2	330	9 AA763276	AA763276 vv89d04.r
37	315.4	40.5	480	10 BE112018	BE112018 UI-R-BJ1-
38	312.4	40.2	554	12 BF555975	BF555975 UI-R-A1-e
39	311.8	40.1	909	12 BF790243	BF790243 602249777
40	309	39.7	614	12 BG223758	BG223758 IM0001860
41	300.6	38.6	498	9 A1171376	A1171376 E87217332
42	299.2	38.5	454	9 AA924895	AA924895 UI-R-A1-e
43	298.2	38.3	446	13 B1293509	B1293509 UI-R-DK0-
44	298.2	38.3	451	12 BG375717	BG375717 UI-R-CS0-
45	297.6	38.3	445	12 BF394985	BF394985 UI-R-CM0-

ALIGNMENTS

RESULT 1	AK003105	934 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus adult male heart cDNA, RIKEN full-length enriched library, clone:1010001C09:small muscle protein, X-linked, full insert sequence.				
ACCESSION	AK003105				
VERSION	AK003105.1	GI:12833554			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male heart cDNA to mRNA, clone lib:RIKEN full-length enriched mouse cDNA library clone:1010001C09.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	10349636				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
PUBMED	20493374				
PUBMED	11042159				

TITLE
MEDLINE
PUBMED
REFERENCE
AUTHORS

Yoshihide Hayashizaki, Y.
Toyo-Oka, K., Wang, K.H., Melz, C., Whitteker, C., Wilming, L.,
Myshaw-Boris, A., Yoshida, K., Haegawa, Y., Kawaji, H., Kohzuki, S.
and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection
Nucleic Acids Res. 29(18):6821-6830 (2001)

11217851
21085660

(bases 1 to 934)
Aadachi, J., Adachi, T., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bull, C.,
Carninci, P., Fukuda, S., Fukuuchi, Y., Furuno, M., Hangaki, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirose, T., Horii, F.,
Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kaenaka, T.,
Kato, H., Kawajiri, K., Kojima, Y., Komori, H., Kouda, M., Koyama, K.,
Kuribayashi, C., Matsuyama, T., Miyazaki, A., Nishikawa, K.,
Numataki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schraml, L., Shibata, K., Shibata, Y., Shingawa, A., Takahashi, F.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Tanaka, T.,
Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamaneke, T.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

Direct Submission
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 220-0045, Japan [E-mail: genome-res@gs.c.riken.go.jp],
URL: http://genome.gs.c.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216

Please visit our web site (<http://genome.gs.c.riken.go.jp/>) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAAGATCCAGACCTCTTTTCTTTTTTTTNN 3'], cDNA was
prepared by using triphase thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 25.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'-
GAGAGAAGATTCGAGTTATTAAATTATCCCCCCCCCCC 3']. cDNA was cleaved
with XhoI and SacI. Cloning sites, 5' end: XhoI; 3' end: SacI.
Host: SOLR.

Location/Qualifiers

FEATURES
source

1..934
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="RANTOM:DB:2310075E21"
/db_xref="MGD:MGI:1893249"
/db_xref="taxon:10090"
/clone="2310075E21"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

gene
misc_feature

1..934
/gene="Smx"
/note="data source:MGD, source key:MGI:1913356,
evidence:ISS
Small muscle protein, X-linked"
/db_xref="MGD:MGI:1913356"
BASE COUNT 276 a 186 c 211 g 261 t

ORIGIN

Query Match Score 770; DB 11; Length 934;
Best Local Similarity 99.4%; Pred. No. 2e-182;
Matches 773; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 117 TTTGTGTTTGTGACCTGGCCGCTGGAGCTGTCTCTCAGGCACTAAACCAATCCAGAGA 176
 DB 81 TTTGTGTTTGTGACCTGGCCGCTGGAGCTGTCTCTCAGGCACTAAACCAATCCAGAGA 140
 OY 177 GCAGGGGCTAAGACCTTGTGAATATGTGAAAGCAGCCAAATTTCCAACTCAGACCAATCCA 236
 DB 141 GCAGGGGCTAAGACCTTGTGAATATGTGAAAGCAGCCAAATTTCCAACTCAGACCAATCCA 200
 OY 237 GGGCAATATCAATATTCATTCAGAGGAGCTTTCCTCCGAGAGCTGGAGCCCTCCAGAAAG 296
 DB 201 GGGCAATATCAATATTCATTCAGAGGAGCTTTCCTCCGAGAGCTGGAGCCCTCCAGAAAG 260
 OY 297 GAAAGAGAGTACTCTCTGAACTAGAGAGAGGAGCTCTTACCACTCAGAGGAAAGAAAGCC 356
 DB 261 GAAAGAGAGTACTCTCTGAACTAGAGAGAGGAGCTCTTACCACTCAGAGGAAAGAAAGCC 320
 OY 357 AATTCTGGAATGAAGAAATTTCCAGAGCTGTGTGCAACTGTCTGAGATCCAAATATGT 416
 DB 321 AATTCTGGAATGAAGAAATTTCCAGAGCTGTGTGCAACTGTCTGAGATCCAAATATGT 380
 OY 417 TAAAGTGACTGAATTTGTCTCCCAAGGTGAACAGTATCGAAAGGACACAAAGTTTC 476
 DB 381 TAAAGTGACTGAATTTGTCTCCCAAGGTGAACAGTATCGAAAGGACACAAAGTTTC 440
 OY 477 ACATTGATGCTTGAATCAGAGATGCAATTTCTGTGACGTGTCTTTCAGAGGAGGAGAAA 536
 DB 441 ACATTGATGCTTGAATCAGAGATGCAATTTCTGTGACGTGTCTTTCAGAGGAGGAGAAA 500
 OY 537 AACATGGGTGAATTAACCACTCTGAACTTTTATACATTTTATGATGATCAAA 596
 DB 501 AACATGGGTGAATTAACCACTCTGAACTTTTATACATTTTATGATGATCAAA 560
 OY 597 ACCCTCTGAATGCCCAAGACTCTAGCAAAATATCTGTGTGAATTTATTTCTTCC 656
 DB 561 ACCCTCTGAATGCCCAAGACTCTAGCAAAATATCTGTGTGAATTTATTTCTTCC 620
 OY 657 TTTTACTGTTGATGATTTTCTCACTTATGATGATTTTGGCACTTGTAGAGCAATCAG 716
 DB 621 TTTTACTGTTGATGATTTTCTCACTTATGATGATTTTGGCACTTGTAGAGCAATCAG 680
 OY 717 CACAGCAATTTACCACTGGAGAGTGTGTTTGAAGAGATGATTTTATGAAGG 776
 DB 681 CACAGCAATTTACCACTGGAGAGTGTGTTTGAAGAGATGATTTTATGAAGG 740
 OY 777 GG 778
 DB 741 GG 742
 RESULT 4
 B0554133 613 bp mRNA linear EST 20-JUN-2002
 LOCUS H4026C08-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
 DEFINITION B0554133
 H4026C08 5', mRNA sequence.
 ACCESSION B0554133
 VERSION B0554133.1 GI:21455021
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranial; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 613)
 VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin,
 P.R., Stagg, C.A., Bassey, U., Alida, K., Hamatani, T., Kargul, G.J.,
 Luo, A.G. and Ko, M.S.H.
 Assembly, verification, and initial annotation of NIA 7.4K mouse
 cDNA clone set
 Unpublished (2002)
 Other ESTs: H4026C08-3
 Contact: Yong Qian
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
 Email: cdna@lsun.gic.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit http://lsun.gic.nia.nih.gov/cDNA/NIA_7_4k.html for details.
 Plate: H4026 row: C column: 08
 Seq primer: -21M13 Reverse
 High quality sequence stop: 613
 POLY-A=No.

FEATURES
 SOURCE
 Location/Qualifiers
 1..613
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="niaEST:H4026C08-5"
 /db_xref="taxon:10090"
 /clone="H4026C08"
 /clone_1bp="NIA Mouse 7.4K cDNA Clone Set"
 /sex="mixed"
 /dev_stage="mixed"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
 clone is among a rearrayed set of 7,407 clones from more
 than 20 cDNA libraries."

BASE COUNT 191 a 140 c 138 g 144 t
 ORIGIN
 Query Match 78.2%; Score 608.4; DB 14; Length 613;
 Best Local Similarity 99.8%; Pred. No. 6,1e-142;
 Matches 609; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 22 CAGAGCATCTCAGCTATTTTACGCCATGAAAGAGCTGGAATGATGCCCGCTCAGA 81
 DB 4 CCGAGCATCTCAGCTATTTTACGCCATGAAAGAGCTGGAATGATGCCCGCTCAGA 63
 OY 82 GACACCGGAGTCTCTTATCTGTAAAGCTTTTGTGTTTGTGACCTGGCCGC 141
 DB 64 GACACCGGAGTCTCTTATCTGTAAAGCTTTTGTGTTTGTGACCTGGCCGC 123
 OY 142 TGGGACTGTCTCAGGAGTAAACCAATCCAGAGCAGGGCTAAGACTTGTGAATATG 201
 DB 124 TGGGACTGTCTCAGGAGTAAACCAATCCAGAGCAGGGCTAAGACTTGTGAATATG 183
 OY 202 TCGAAGAGCAATTTTCCAAAGTCAAGGCAATCCAGGCAATATCAATTTCAATGGGA 261
 DB 184 TCGAAGAGCAATTTTCCAAAGTCAAGGCAATCCAGGCAATATCAATTTCAATGGGA 243
 OY 262 GCTTTGTCGCGGAGTGTGGCAGCTCCGAGAGGAAAGAGTACTCTGAAACTGAG 321
 DB 244 GCTTTGTCGCGGAGTGTGGCAGCTCCGAGAGGAAAGAGTACTCTGAAACTGAG 303
 OY 322 GAGGAGCTCTTACCACTCAGAGGAAAGAGCAATTTCTGGAATGAAGAAATTTCCA 381
 DB 304 GAGGAGCTCTTACCACTCAGAGGAAAGAGCAATTTCTGGAATGAAGAAATTTCCA 363
 OY 382 GGACCTGTGTCACCTTGTGATGATCCAAATATGTTAAAGTGAATGAAATTTGCCCC 441
 DB 364 GGACCTGTGTCACCTTGTGATGATCCAAATATGTTAAAGTGAATGAAATTTGCCCC 423
 OY 442 AAAGTGAAAGTATGTAAGGACACAAAGTTCACTGATGATGCTTGAATCAGAGGA 501
 DB 424 AAAGTGAAAGTATGTAAGGACACAAAGTTCACTGATGATGCTTGAATCAGAGGA 483
 OY 502 TGCATTGCTGAGTGTGTTTTCAGAGGAGAAACAAATGAGTTGAATTAACAATTT 561
 DB 484 TGCATTGCTGAGTGTGTTTTCAGAGGAGAAACAAATGAGTTGAATTAACAATTT 543
 OY 562 CCGTAACATTTTATACATTTTATGATGATCACAACCTCTGTAATGCCAAGACTTAG 621
 DB 544 CCGTAACATTTTATACATTTTATGATGATCACAACCTCTGTAATGCCAAGACTTAG 603
 OY 622 CAAAATATTC 631
 DB 604 CAAAATATTC 613

RESULT 5
AA033164
LOCUS
DEFINITION m37c11.1 Soares mouse embryo NMME13.5 14.5 Mus musculus cDNA
VERSION AA033164
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 504)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Gaisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepcie, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Maria M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LIND, contact the
IMAGE Consortium (info@image.lind.gov) for further information.
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 494.
Location/Qualifiers
1. 504
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:465716"
/clone_lib="Soares mouse embryo NMME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGGAAATTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2 1; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Patricia Bonaldo."

BASE COUNT 152 a 113 c 126 g 113 t
ORIGIN

Query Match 64.8%; Score 504; DB 9; Length 504;
Best Local Similarity 100.0%; Pred. No. 8.9e-116;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 GCCACATGAAGAAGCATGATCCCGCTCAGAGACACCGGAGTTCCTCTA 102
DB 1 GCCACATGAAGAAGCATGATCCCGCTCAGAGACACCGGAGTTCCTCTA 60
QY 103 TCTGTAAAGGCTTTTGTGTTTGTGACCTGGCGGCTGGAGCTCTCTCAGGAGTA 162
DB 61 TCTGTAAAGGCTTTTGTGTTTGTGACCTGGCGGCTGGAGCTCTCTCAGGAGTA 120
QY 163 AACCAATCCAGAGGCGCTTAAGACTTGTGATATGTGGAAGCAGCAATTCCTAAC 222

DB 121 AACCAATCCAGAGGCGCTTAAGACTTGTGATATGTGGAAGCAGCAATTCCTAAC 180
QY 223 GTCAGAGCCATCCAGCGCAATTCATATTCATGAGGAGCTTGTCTCCGAGAGCTGG 282
DB 181 GTCAGAGCCATCCAGCGCAATTCATATTCATGAGGAGCTTGTCTCCGAGAGCTGG 240
QY 283 CAGCTCCAGAGGAGAGAGTACTCTGAACTGAGGAGGAGCTCTTCACTCA 342
DB 241 CAGCTCCAGAGGAGAGAGTACTCTGAACTGAGGAGGAGCTCTTCACTCA 300
QY 343 CAGGAAAAGAGCCATTCCTGGAATGAGAAATTCAGAGCCTGTTGCACTTGTCT 402
DB 301 CAGGAAAAGAGCCATTCCTGGAATGAGAAATTCAGAGCCTGTTGCACTTGTCT 360
QY 403 CAGATCCAAAATGTTAAAGGAACTGAAATTTGTCCCAAGGTGAACGTGCGAAA 462
DB 361 CAGATCCAAAATGTTAAAGGAACTGAAATTTGTCCCAAGGTGAACGTGCGAAA 420
QY 463 GGCACAAAAGTTCAATGATGATCTTGAATCAGAGATGCAATTCCTGAGCTGTTT 522
DB 421 GGCACAAAAGTTCAATGATGATCTTGAATCAGAGATGCAATTCCTGAGCTGTTT 480
QY 523 TCCAAAGGAGAAAACAAATGCGT 546
DB 481 TCCAAAGGAGAAAACAAATGCGT 504

RESULT 6
B0176246/c
LOCUS
DEFINITION B0176246 657 bp mRNA linear EST 30-APR-2002
UI-M-DJ2-bwg-d-13-0-UI.81 NIH BMAP DJ2 Mus musculus cDNA clone
UI-M-DJ2-bwg-d-13-0-UI 3', mRNA sequence.
B0176246
B0176246.1 GI:20351738
EST.
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 657)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL MEDLINE
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mbs@nhi.nih.gov
Tissue Procurement: Dr. Robin Davison
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-40, >At rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=yes
FEATURES
source
1. 657
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="UI-M-DJ2-bwg-d-13-0-UI"
/clone_lib="NIH_BMAP_DJ2"
/tissue_type="subfornical organ and postrema"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: brain; Vector: pT73-Pac (Pharmacia) with a

QY 121 GTGTTTTCGACCTGGCCGCTGGGACTGT-CCTCAGGAGTAAACCAATCCAGAGACA 179
 Db 496 G-GTTTTCGACCTGGCCGCTGGGACTGTCTCAGCAGTAAACCAATCCAGAGACA 438
 QY 180 GGGCTAAGACCTTGTGATATGTGCAAGCAGCAATTTCCAACTC-AGAGCCATCCAGG 238
 Db 437 GGGCTAAGACCTTGTGATATGTGCAAGCAGCAATTTCCAACTC-AGAGCCATCCAGG 378
 QY 239 CGAATATCAATATTCGAATGGAGCCTTTCCTCCGGAGCTGGGACGCTCCCAAGAGA 298
 Db 377 CGAATATCAATATTCGAATGGAGCCTTTCCTCCGGAGCTGGGACGCTCCCAAGAGA 318
 QY 299 AAGAGAGTACCTCGAAACCTGAGGAGGAGCTCCACCACTCAGAGAGAAAGAGCAA 358
 Db 317 AAGAGAGTACCTCGAAACCTGAGGAGGAGCTCCACCACTCAGAGAGAAAGAGCAA 258
 QY 359 TTCTGGAATGAGAAATTTCCAGGACCTGTGTCACTTGTCTGAGATCCAAATGTTA 418
 Db 257 TTCTGGAATGAGAAATTTCCAGGACCTGTGTCACTTGTCTGAGATCCAAATGTTA 198
 QY 419 AAGTGAATGAAATTTGTCCCAAGGTGAACAGTATCGAAAGAGACAAAAGTTTAC 478
 Db 197 AAGTGAATGAAATTTGTCCCAAGGTGAACAGTATCGAAAGAGACAAAAGTTTAC 138
 QY 479 ATTGATGCTTGAATGATGAGAGATGATTTGTTGAGCGTGTTCGAAAGGAGAAAAA 538
 Db 137 ATTGATGCTTGAATGATGAGAGATGATTTGTTGAGCGTGTTCGAAAGGAGAAAAA 78
 QY 539 CAATGGTTGAATTAACAATCTTCTGAAACATTTTATACA 578
 Db 77 CAATGGTTGAATTAACAATCTTCTGAAACATTTTAAAAA 38

RESULT 9 482 bp mRNA linear EST 29-MAY-1997
 LOCUS AA434782
 DEFINITION ve33c01.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:818976
 ACCESSION AA434782
 VERSION AA434782.1 GI:2139696
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 482)
 Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepcoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HMNI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Maier M/Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:487256
 Seq primer: -28ml3 rev2 ET from Amerham
 High quality sequence Etop: 411.
 FEATURES
 SOURCE
 1..482
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:818976"
 /clone_1ib="Soares mouse NBMH"

/sex="male"
 /tissue_type="heart"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'] TGTATCAATCTGAAAGTGGAGCGCCGCAAGATTTTTTTTTTTTTTTTTT [3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 141 a 118 c 124 g 99 t
 ORIGIN

Query Match 58.3%; Score 453.4; DB 9; Length 482;
 Best Local Similarity 99.6%; Pred. No. 4,3e-103;
 Matches 465; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GCTCTGAGACTGGAGAGACAGACACTCCAGCTATTTCAGCCATGAAAGCACTG 60
 Db 17 GCTCTGAGACTGGAGAGACAGACACTCCAGCTATTTCAGCCATGAAAGCACTG 76
 QY 61 GAATGAGATCCCGCTCAGAGACACCGGAGTTCTTATCTGTAAGCGCTTTT 120
 Db 77 GAATGAGATCCCGCTCAGAGACACCGGAGTTCTTATCTGTAAGCGCTTTT 136
 QY 121 GTGTTTTCGACCTGGCCGCTGGGACTGTCTCAGCAGTAAACCAATCCAGAGACA 180
 Db 137 GTGTTTTCGACCTGGCCGCTGGGACTGTCTCAGCAGTAAACCAATCCAGAGACA 196
 QY 181 GGGCTAAGACCTTGTGATATGTGCAAGCAGCAATTTCCAACTC-AGAGCCATCCAGG 240
 Db 197 GGGCTAAGACCTTGTGATATGTGCAAGCAGCAATTTCCAACTC-AGAGCCATCCAGG 256
 QY 241 AATATCAATATTCGAATGGAGCCTTTCCTCCGGAGCTGGGACGCTCCCAAGAGAA 300
 Db 257 AATATCAATATTCGAATGGAGCCTTTCCTCCGGAGCTGGGACGCTCCCAAGAGAA 316
 QY 301 GAGAGTACTCCTGAACTGAGAGGAGCTCTTACCACTCAGAGAGAAAGAGCCAAAT 360
 Db 317 GAGAGTACTCCTGAACTGAGAGGAGCTCTTACCACTCAGAGAGAAAGAGCCAAAT 375
 QY 361 CCTGGAATGAGAAATTTCCAGGACCTGTGTCACTTGTCTGAGATCCAAATGTTAA 420
 Db 376 CCTGGAATGAGAAATTTCCAGGACCTGTGTGTCACTTGTCTGAGATCCAAATGTTAA 435
 QY 421 AGTGAATGAAATTTGTCCCAAGGTGAACAGTATCGAAAGAGACA 467
 Db 436 AGTGAATGAAATTTGTCCCAAGGTGAACAGTATCGAAAGAGACA 482

RESULT 10 486 bp mRNA linear EST 26-JUN-1998
 LOCUS A1035961
 DEFINITION ub50b12.r1 Soares mammary gland NBMMG Mus musculus cDNA clone IMAGE:1381151 5' similar to TR:Q93031 Q93031 COSMID U228D4 ;, mRNA
 ACCESSION A1035961
 VERSION A1035961.1 GI:3259652
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 486)
 Maier, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Stepcoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 REFERENCE
 AUTHORS

100

100

Db 338 GATCACAACCTCTGATGCCAAGACTGACGAAAAATCTCTGTTTGACATTATA 279
 Qy 649 TTTCTCTTCTTACTGTTGATTCATTTCTGACTTGTAGATTGTTGGACCTGTAGAG 708
 Db 278 TTTCTCTTCTTACTGTTGATTCATTTCTGACTTGTAGATTGTTGGACCTGTAGAG 219
 Qy 709 CAATCAGCAGACGATTTACCACTGGAGAGTGTGTTTGAGAGAGATGTGATTTT 768
 Db 218 CAATCAGCAGACGATTTACCACTGGAGAGTGTGTTTGAGAGAGATGTGATTTT 159
 Qy 769 ATGAAGGCGG 778
 Db 158 ATGAAGGCGG 149

RESULT 12
 AA060214 490 bp mRNA linear EST 23-SEP-1996
 LOCUS mb65h06.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

DEFINITION IMAGE:481019 5', mRNA sequence.

ACCESSION AA060214
 VERSION AA060214.1 GI:1553903

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 490)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

FEATURES
 source
 1. 490
 /db_xref="taxon:10090"
 /clone="IMAGE:481019"
 /clone_lib="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pRT3D (pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT3D vector (pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."

BASE COUNT 152 a 110 c 124 g 104 t

Query Match 54.7%; Score 425.4; DB 9; Length 490;
 Best Local Similarity 99.8%; Pred. No. 4.5e-96;
 Matches 426; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 47 CATGAAGACCTGATGATTCCTCCGCTCAGAGACACCGGAGTTCCTTATCT 106
 Db 15 CATGAAGACCTGATGATTCCTCCGCTCAGAGACACCGGAGTTCCTTATCT 74
 Qy 107 GTAAAGCGCTTTTGTGTTTGTGTTGACCTGCGGCTGTGCTCTCAGGAGTAAAC 166
 Db 75 GTAAAGCGCTTTTGTGTTTGTGTTGACCTGCGGCTGTGCTCTCAGGAGTAAAC 134
 Qy 167 AATCCAGAGAGAGGCTAAGACCTTGATATGTCGAGAGAGCAATTTTCAACGTCA 226
 Db 135 AATCCAGAGAGAGGCTAAGACCTTGATATGTCGAGAGAGCAATTTTCAACGTCA 194
 Qy 227 GAGCCATCCAGGCGAATATCAATATTCATGAGAGCTTTGTCGAGAGCTGGCAGC 286
 Db 195 GAGCCATCCAGGCGAATATCAATATTCATGAGAGCTTTGTCGAGAGCTGGCAGC 254
 Qy 287 CTCCGAGAGAGAGAGAGTATCTCTGTTGAACTGAGAGAGAGAGTCTTACCCTCAGAG 346
 Db 255 CTCCGAGAGAGAGAGAGTATCTCTGTTGAACTGAGAGAGAGAGTCTTACCCTCAGAG 314
 Qy 347 AAAAGAGCCATTTCTGGAATGAGAAATTTCCAGAGCTGTGTCATCTGTGAGA 406
 Db 315 AAAAGAGCCATTTCTGGAATGAGAAATTTCCAGAGCTGTGTCATCTGTGAGA 374
 Qy 407 TCCAAATGTTAAAGTGAATGAAATTTGTCCTCCAAAGGTGAACATGCGAAAGAC 466
 Db 375 TCCAAATGTTAAAGTGAATGAAATTTGTCCTCCAAAGGTGAACATGCGAAAGAC 434
 Qy 467 ACAAAAG 473
 Db 435 ACAAAAG 441

RESULT 13
 W36988 436 bp mRNA linear EST 11-SEP-1996
 LOCUS mb65h11.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone

DEFINITION IMAGE:334269 5', mRNA sequence.

ACCESSION W36988
 VERSION W36988.1 GI:1318245

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 436)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Marra M/Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu

FEATURES
 source
 1. 436
 /db_xref="taxon:10090"
 /clone="IMAGE:334269"
 /clone_lib="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"

High quality sequence stop: 436.
 Seg primer: ETPRimer
 Location/Qualifiers

IMAGE:331102.5', mRNA sequence.
 W13738
 W13738.1 GI:1287926
 EST.
 SOURCE
 ORGANISM

REFERENCE
AUTHORS

house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 448)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellander, K., Stepec, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:212502

TITLE
 JOURNAL
 COMMENT

Seq primer: ETPPrimer
 High quality sequence stop: 441.
 Location/Qualifiers

FEATURES
SOURCE

1..448
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:331102"
 /clone_1db="Soares mouse p3NMP19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAGAGGAGGAGCGCGGCGGATTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldi. RNA was kindly provided by
 Dr. Minoru Ko (Wayne State University)."
 BASE COUNT 128 a 108 c 113 g 98 t 1 others
 ORIGIN

Query Match 51.6%; Score 401.6; DB 14; Length 448;
 Best Local Similarity 96.9%; Pred. No. 4.1e-90;
 Matches 431; Conservative 0; Mismatches 10; Indels 4; Gaps 2;

QY 1 GCTCTCAGAGCTGAGAGAGACAGCACTCCAGCTATTTCAGCCATGAAAGCACTG 60
 DB 1 GCTCTCAGAGCTGAGAGAGACAGCACTCCAGCTATTTCAGCCATGAAAGCACTG 60
 QY 61 GAATTGAGATCCCGCTCAGAGAGACACCGGAGTCTTCTATCTCTGTAAGCGCTTTT 120
 DB 61 GAATTGAGATCCCGCTCAGAGAGACACCGGAGTCTTCTATCTCTGTAAGCGCTTTT 120
 QY 121 GTGTTTTCAGACTGCGCGCTGAGAGTGTCTCTCAGAGAGTAAACCAATCCAGAGAGAG 180
 DB 121 GTGTTTTCAGACTGCGCGCTGAGAGTGTCTCTCAGAGAGTAAACCAATCCAGAGAGAG 180
 QY 181 GGCTTAAGACTTGTGATATGTCGAAGACCAATTTCCAACTGAGAGCCATCCAGGCG 240
 DB 181 GGCTTAAGACTTGTGATATGTCGAAGACCAATTTCCAACTGAGAGCCATCCAGGCG 240
 QY 241 AATATCATATATTCGAATGGAGCGCTTGTCCGGAGAGCTGGGCAAGGAGGAAA 300
 DB 241 AATATCATATATTCGAATGGAGCGCTTGTCCGGAGAGCTGGGCAAGGAGGAAA 300

QY 301 GAGAGTACTCTGAAACTGAGAGAGAGCTCTACCACTCAGAGAGAAAGAACCAATT 360
 DB 301 GAGAGTACTCTGAAACTGAGAGAGAGCT-CTACCACTCAGAGAGAAAGAACCAATT 359
 QY 361 CTTGAAATGAAGAAATTTCCAGAGACCTGTTGTCAACTTGTCTGAGATCCAAATGTTTAA 420
 DB 360 CTTGAAATGAAGAAATTTCCAGAGACCTGTTGTCAACTTGTCTGAGATCCAAATGTTTAA 419
 QY 421 AGTGAATGAATTTGTCCCAAG 445
 DB 420 TG--ACTGAATTTGTCCCAAG 441

Search completed: April 15, 2003, 22:02:21
 Job time : 1620.62 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 ; Search time 1762.69 Seconds

(without alignments)
12845.099 Million cell updates/sec

Title: US-09-647-019-1

Perfect score: 778
Sequence: 1 gctctcagactgagagag.....tgctattttatgaagggg 778

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	778	100.0	787	10	AY026524 Mus muscu
2	777	99.9	943	10	MMU245772 Mus muscu
3	774.8	99.6	936	10	AF364070 Mus muscu
4	552.8	71.1	892	10	AF364071 Mus muscu
5	389.8	50.1	885	9	HS2450584 Mus muscu
6	389.8	50.1	886	9	AF123505 Mus muscu
7	388.2	49.9	886	6	AX322774 Mus muscu
8	349	44.9	835	9	BC005948 Mus muscu
9	270	34.7	228031	2	AL732396 Mus muscu
10	215.8	27.7	188670	2	AC127606 Mus muscu
11	214.2	27.5	82586	2	AC096040 Mus muscu
12	181	23.3	239667	2	AL731811 Mus muscu
13	162.4	20.9	587	6	AX332418 Mus muscu
14	146.8	18.9	36503	9	U73508 Mus muscu
15	146.8	18.9	124186	9	AL772392 Mus muscu
16	114.6	14.7	923	5	AF343894 Mus muscu
17	109.6	14.1	290	6	AX322783 Mus muscu
18	98	12.6	47440	9	U73509 Mus muscu
19	98	12.6	150319	2	AL772370 Mus muscu
20	82	10.5	82586	2	AC096040 Mus muscu
21	53.4	6.9	107328	2	AL160057 Mus muscu
22	48.6	6.2	43470	9	AP001216 Mus muscu
23	48.6	6.2	211201	9	AC009958 Mus muscu
24	46.2	5.9	146403	2	AC027067 Mus muscu
25	45.6	5.9	191676	2	AC105075 Mus muscu
26	44.4	5.7	1141	6	AX083744 Mus muscu
27	43.4	5.6	7218	6	I66494 Sequence 14
28	43	5.5	185571	9	AL390876 Human DNA
29	43	5.5	247850	10	AJ421478 Human DNA
30	42.8	5.5	1981	3	BOL27835 Human DNA
31	42.8	5.5	141935	2	AC094920 Rattus no
32	42.6	5.5	131221	2	AC103601 Mus muscu
33	42.6	5.5	133897	2	AC112138 Homo sapi
34	42.6	5.5	154350	2	AC016948 Homo sapi
35	42.4	5.4	171750	2	AC016845 Homo sapi
36	42.4	5.4	181994	2	AC090822 Homo sapi
37	42.4	5.4	184925	2	AC027608 Homo sapi
38	42.4	5.4	189231	9	AC090138 Homo sapi
39	42.4	5.4	197103	9	AL590036 Human DNA
40	42.2	5.4	179788	2	AL645525 Mus muscu
41	42.2	5.4	219788	2	AL645525 Mus muscu
42	42.2	5.4	241701	2	AC123583 Mus muscu
43	42	5.4	1571	9	AK021863 Homo sapi
44	41.6	5.3	154594	9	AL161722 Human DNA
45	41.4	5.3	107819	9	AC010465 Homo sapi

ALIGNMENTS

RESULT 1
AY026524 787 bp mRNA linear ROD 28-JUN-2001
LOCUS AY026524 Mus musculus muscle-specific protein CSL (Csl) mRNA, complete cds.
DEFINITION AY026524
ACCESSION AY026524 GI:14575061
VERSION
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 787)
REFERENCE
AUTHORS Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C., Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F., Mohun,T. and Harvey,R.P.

OY		943 bp	MWU245772	RNA linear ROD 12-APR-2001
LOCUS			MMU245772	
DEFINITION			Mus musculus mRNA for stretch responsive muscle (X-chromosome)	
ACCESSION			AJ245772	
VERSION			AJ245772.1 GI:10178962	
KEYWORDS			Srnx gene; stretch responsive muscle (X-chromosome).	
SOURCE			house mouse.	
ORGANISM			Mus musculus	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 943) Kemp,T.J., Sadusky,T.J., Simon,M., Brown,R., Eastwood,M., Sassoon,D.A. and Coulton,G.R. Identification of a novel stretch-responsive skeletal muscle gene (Smpx)	
TITLE	JOURNAL	Genomics 72 (3), 260-271 (2001)		
PUBLISHED	MEDLINE	21285047		
REFERENCE	FUNDED	11401441		
AUTHORS		2 (bases 1 to 943) Kemp,T.J.		
JOURNAL	TITLE	Direct Submission		
		Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial College School of Medicine, SAF Building, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM		
FEATURES	source	Location/Qualifiers		
		1..943	/organism="Mus musculus"	
			/strain="C57BL/10"	
			/db_xref="taxon:10090"	
			/tissue_type="skeletal muscle"	
	gene	1..943	/gene="Srinx"	
		1..214	/gene="Srinx"	
		215...472	/gene="Srinx"	
	5' UTR		/codon_start=1	
			/product="stretch responsive muscle (X-chromosome)"	
			/protein_id="CAC08493.1"	
			/db_xref="GI:10178963"	
			/translation="MSKPISVRAIOAININIPGAFFPGAGQPPRKESPTPEEGA PTTSEKKPPIPMGKKFPVPVMLSIQVKSELKFVRKEGR"	
		473..943	/feature="SRNX"	
			/gene="Srinx"	
		535..539	/gene="Srinx"	
	misc_feature		/function="mRNA destabilising motif"	
		585..590	/gene="Srinx"	
			/function="mRNA destabilising motif"	
		658..662	/gene="Srinx"	
	misc_feature		/function="mRNA destabilizing motif"	


```

QY 361 CCGTGAATGAGAAATTTCCAGACCTGTGTCTCACTTGTCTGATCCAAATGTTAA 420
Db 361 CCGTGAATGAGAAATTTCCAGACCTGTGTCTCACTTGTCTGATCCAAATGTTAA 420
QY 421 AGTGAAGTGAATTTTCCCAAGGTGAACAGTACGAGAAAGACCAAAAGTTCAAT 480
Db 421 AGTGAAGTGAATTTTCCCAAGGTGAACAGTACGAGAAAGACCAAAAGTTCAAT 480
QY 481 TGGATGCTTGAATTCAGAGATGATTCATTCGTTGACGTTTTCGAAGGAGAAAAACA 540
Db 481 TGGATGCTTGAATTCAGAGATGATTCATTCGTTGACGTTTTCGAAGGAGAAAAACA 540
QY 541 ATGGTGTGAATTAACAATCTCTGAACATTTATATATATATATATATATATATATAT 600
Db 541 ATGGTGTGAATTAACAATCTCTGAACATTTATATATATATATATATATATATATAT 600
QY 601 CCTGAATGCCAGACCTTACGAAATAATCTGTGTGTATATATATATATATATATAT 660
Db 601 CCTGAATGCCAGACCTTACGAAATAATCTGTGTGTATATATATATATATATATAT 660
QY 661 ACTGTGTGATTTCTCACTTTAGCTTACATTTTGGACCTGTGTAGAGCAAAATCAGACA 720
Db 661 ACTGTGTGATTTCTCACTTTAGCTTACATTTTGGACCTGTGTAGAGCAAAATCAGACA 720
QY 721 CGAATTTACAACTGGGAGTGTGTTTGGAGAGATGATTTTATGAAAGGGGG 778
Db 721 CGAATTTACAACTGGGAGTGTGTTTGGAGAGATGATTTTATGAAAGGGGG 778

```

```

RESULT 4
AF364071 892 bp mRNA linear ROD 04-MAY-2001
LOCUS Rattus norvegicus Smx protein (Smx) mRNA, complete cds.
ACCESSION AF364071.1 GI:13940509
VERSION AF364071.1 GI:13940509
KEYWORDS
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE 1 (bases 1 to 892)
AUTHORS Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
TITLE Identification, mapping, and genomic structure of a novel
X-chromosomal human gene (Smx) encoding a small muscular protein
JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
MEDLINE 20065879
PubMed 10598820
REFERENCE 2 (bases 1 to 892)
AUTHORS Patzak,D.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
Fleischmannstr. 42-44, D-17487 Greifswald, Germany
FEATURES
source
1..892
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
<1..892
/gene="Smx"
190..447
/gene="Smx"
/codon_start=1
/product="Smx protein"
/protein_id="AA50399.1"
/db_xref="GI:13940510"
/translational="MSKOPTSNVRSIQANINIPWGAFFPGAPPRKESTPGTAKGA
PATPEKKPVPMKKFPGPVNLSIQNVSEIKYVKRGQ"
504..509
/gene="Smx"
/evidence=experimental
862..867
/gene="Smx"

```

```

polya_signal
polya_signal
polya_signal

```

```

Query Match 71.1%; Score 552.8; DB 10; Length 892;
Best Local Similarity 86.8%; Pred. No. 8.3e-135;
Matches 670; Conservative 0; Mismatches 62; Indels 40; Gaps 4;

```

```

BASE COUNT 271 a 183 c 193 g 245 t
ORIGIN
8
GGAGTGGAGAGAGACAGACGCTCCAGCTATTTTCAAGCCATGTAAGAAAGCACTGGAATTGA 67
1
GGACCGAGAGAGAGACAGACGCTCCAGCTATTTTCAAGCCATGTAAGAAAGCACTGGAATTGA 60
QY 68 GATCCCGCTCAGAGAGACAGCGGAGTCTTCTATCTGTAAGAGCGCTTTTGTGTTT 127
Db 61 GATCTCAGCTCAGAGAGACAGCGGAGTCTTCTATCTGTAAGAGCG--TTTCTGTTT 118
QY 128 TGCACCTGCGCGCTGGAGCTGTCTCAGGCAATTAACCAATCCAGAGAGAGGCTAAG 187
Db 119 TGCACCTGCGCGCTGGAGCTGTCTCAGGCAATTAACCAATCCAGAGAGAGGATTAAG 178
QY 188 ACCTGGAATATGTCAGAGAGAGCAATTTCCAGAGTCAAGCCATCCAGGCAATATCA 247
Db 179 ACCTTGAATATGTCAGAGAGAGCAATTTCCAGAGTCAAGCCATCCAGGCAATATTA 238
QY 248 ATATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 307
Db 239 ATATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298
QY 308 CTCCTGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
Db 299 CCCCCTGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 358
QY 368 TGAAGAAATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
Db 359 TGAAGAAATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 418
QY 428 TGAATTTGTCGCCAAGAGTGAACAGTATGCAAGAGACCAAAAGTTCAATGATGC 487
Db 419 TAAATATGTCGCCAAGAGTGAACAGTATGCAAGAGAGAGAGAGAGAGAGAGAG 466
QY 488 TTGAATCAGAGAGATGATTCGTTGACGTTTTCGAAGAGAGAGAGAGAGAGAGAG 547
Db 467 -----TTGACGTCGTTTTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 501
QY 548 GAAATTAACAACCTCTGACATTTTATACATTTGATGATGATCAAAACCTCGAAT 607
Db 502 GAAATTAACAACCTCTGACATTTTATATATTTGATGATGATCAAAACCTCGAAT 561
QY 608 GCCCAAGACTCTAGC-AAAAATATCTGTTGTGACATTTATATTTCTCTTACTTGG 666
Db 562 GCCCAAGACTCTAGCAAAAATATCTGTTGTGACATTTATATTTCTCTTACTTGG 621
QY 667 TTGATTTTCACATTTAGCTACATTTTGGACCTGTGAGCAAAATCAGACAGCAAT 726
Db 622 CTGATTTTCACATTTAGCTACATTTTGGACCTGTGAGCAAAATCAGACAGCAAT 681
QY 727 TACAACCTGGAGAGTGTGTTTGGAGAGAGATGATTTTATGAAAGGGGG 778
Db 682 TACAACCTGGAGAGTGTGTTTGGAGAGAGATGATTTTATGAAAGGGGG 733

```

```

RESULT 5
HSA250584 885 bp mRNA linear PRI 12-APR-2001
LOCUS HSA250584 Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
DEFINITION protein (Smx gene).
ACCESSION AJ250584 GI:10178976
VERSION AJ250584.1
KEYWORDS Smx gene; stretch responsive muscle (X-chromosome).
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE 1 (bases 1 to 885)
 AUTHORS Kemp, T.J., Sadusky, T.J., Simon, M., Brown, R., Eastwood, M., Sabsnon, D.A. and Coulton, G.R.
 TITLE Identification of a novel stretch-responsive skeletal muscle gene (Smpx)
 JOURNAL Genomics 72 (3), 260-271 (2001)
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2 (bases 1 to 885)
 AUTHORS Kemp, T.J.
 TITLE Direct Submission
 JOURNAL Submitted (02-NOV-1999) Kemp, T.J., Molecular pathology, Imperial College of Science, Technology and Medicine, SRF Bldg, Level 2, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
 COMMENT Related sequences: AJ245772, U73508 to U73509.
 FEATURES
 source
 1..885
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="skeletal muscle"
 1..885
 /gene="Smpx"
 1..183
 /gene="Smpx"
 184..450
 /gene="Smpx"
 /codon_start=1
 /product="stretch responsive muscle (X-chromosome)"
 /protein_id="CAC08492.1"
 /db_xref="GI:10178977"
 /translation="NMNKSQPVNVRALQANINI PMGAPRGAGPPRRKCTPEVBE GVPTSDSEKKPIPGAKKLPGPAVNLSIQNKSELKVPKAEQ"
 3' UTR
 polyA_signal
 451..885
 /gene="Smpx"
 857..862
 /gene="Smpx"
 172 c 191 g 244 t
 BASE COUNT 278 a 172 c 191 g 244 t
 ORIGIN
 Query Match 50.1%; Score 389.8; DB 9; Length 885;
 Best Local Similarity 73.5%; Pred. No. 7,2e-92;
 Matches 562; Conservative 0; Mismatches 152; Indels 51; Gaps 3;

Db 422 TAAATATGTCCTCCAAAGCTGAACAGTAGTAGAAGAAAAAGATTGATGTGAAGAAAT 481
 Qy 488 TTGAATTCAGAGATGATCTTTCTTGAACGTGTTTTCCAGGAGAAAAACAATGGGT 547
 Db 482 AAAGAGCAGAAAGATGATTCATAGC----- 508
 Qy 548 GAATATAACAATCTCTGAACATTATATCATTTGATATGATGATCAAAACCTCTGAAT 607
 Db 509 -----TCACTAAATTTATATTTATTTGATATGATGATGATGAACTCTGAAT 555
 Qy 608 GCCCAAGACTCTAGCAAAATATCTGTTGATCATTTATATTTCTCTTTACTTGTGT 667
 Db 556 GCTGAGACTCTAGCAAAATGCGCTGTTGATCATTTATATCTCTCTTCTAGTTGGC 615
 Qy 668 TGCATTTCTACTTTTACCTACATTTTGGCACTTGTAGAGCAAAACAGACAGAAATT 727
 Db 616 TGTATTTCTTACTTTATCTTTCATTTTGGCACTTACAGAAACAAATTAAGCCATTAATTC 675
 Qy 728 ACAACCTGGAGAGTGTGTTTGGAGAGATGTGATTTTATGA 772
 Db 676 AACACCTGGAGAGTGTGTTTGGAGAGATGTGATTTTATGA 720
 RESULT 6
 AF129505 886 bp mRNA linear pri 22-DEC-1999
 LOCUS AF129505
 DEFINITION Homo sapiens small muscular protein (SMPX) mRNA, complete cds.
 ACCESSION AF129505
 VERSION AF129505.1 GI:6625646
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 886)
 AUTHORS Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.
 TITLE Identification, mapping, and genomic structure of a novel X-chromosomal human gene (SMPX) encoding a small muscular protein
 JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
 MEDLINE 20065879
 PUBMED 10598820
 REFERENCE 2 (bases 1 to 886)
 AUTHORS Patzak, D.
 TITLE Direct Submission
 JOURNAL Submitted (18-JAN-1999) Molecular Human Genetics, Institut for Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany
 FEATURES
 source
 1..886
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xq22.1-22.2"
 <1..>886
 /gene="SMPX"
 <1..183
 /gene="SMPX"
 184..450
 /gene="SMPX"
 /codon_start=1
 /evidence=not experimental
 /product="small muscular protein"
 /protein_id="AA19343.1"
 /db_xref="GI:6625647"
 /translation="NMNKSQPVNVRALQANINI PMGAPRGAGPPRRKCTPEVBE GVPTSDSEKKPIPGAKKLPGPAVNLSIQNKSELKVPKAEQ"
 190
 /gene="SMPX"
 /note="alternate position for initiation methionine"
 451..>886
 /gene="SMPX"
 857..862
 /gene="SMPX"
 3' UTR
 polyA_signal
 /gene="SMPX"

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5544: contig of 5544 bp in length
 * 5545 5644: gap of 100 bp
 * 5645 11121: contig of 5477 bp in length
 * 11122 11221: gap of 100 bp
 * 11222 23110: contig of 11889 bp in length
 * 23111 23210: gap of 100 bp
 * 23211 42146: contig of 18936 bp in length
 * 42147 42246: gap of 100 bp
 * 42247 74925: contig of 32679 bp in length
 * 74926 75025: gap of 100 bp
 * 75026 153828: contig of 78803 bp in length
 * 153829 153928: gap of 100 bp
 * 153929 172635: contig of 18707 bp in length
 * 172636 172736: gap of 100 bp
 * 172736 228031: contig of 55296 bp in length.

FEATURES
 source
 1. 228031
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone_1fb="RPC1-23"
 /clone_1fb="RPC1-23"
 1. 5544
 /note="assembly fragment:01721
 fragment_chain:1"
 5645. 11121
 /note="assembly fragment:02676
 fragment_chain:1"
 11222. 23110
 /note="assembly fragment:00299
 fragment_chain:1"
 23211. 42146
 /note="assembly fragment:01463
 fragment_chain:1"
 42247. 74925
 /note="assembly fragment:02519
 fragment_chain:1"
 75026. 153828
 /note="assembly fragment:00456
 fragment_chain:1"
 153929. 172635
 /note="assembly fragment:03454
 fragment_chain:2"
 172736. 228031
 /note="assembly fragment:01243
 fragment_chain:2"

BASE COUNT 71213 a 42522 c 42780 g 70814 t 702 others
 ORIGIN

Query Match 34.7%; Score 270; DB 2; Length 228031;
 Best Local Similarity 100.0%; Pred. No. 2.3e-60;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 CGTGCACGCTGTTTCCAAAGGAGAAAAAACAATGCTGTAATAACAACCTCTGAC 568
 DB 187359 CGTGCACGCTGTTTCCAAAGGAGAAAAAACAATGCTGTAATAACAACCTCTGAC 187418

QY 569 ATTATATACATTTGATATGATGATCAAAACCTCTGTAATGCCAAGACTCTAGCAAAAT 628
 DB 187419 ATTATATACATTTGATATGATGATCAAAACCTCTGTAATGCCAAGACTCTAGCAAAAT 187478

QY 629 ATCCGTGTTGATATATATTTCTCTCTTCTTCTGTTGTTGATTTCTCACTTTAGCTAC 688
 DB 187419 ATCCGTGTTGATATATATTTCTCTCTTCTTCTGTTGTTGATTTCTCACTTTAGCTAC 187538

QY 689 ATTTTGGACCTGTTAGAGCAAAATCAGACACGAATTTAACAACCTGGGAAGTGGTTT 748
 DB 187539 ATTTTGGACCTGTTAGAGCAAAATCAGACACGAATTTAACAACCTGGGAAGTGGTTT 187598

QY 749 TGAAGAGATGATGATTTTATGAAGGGGG 778

Db 187599 TGAGAGAGATGATGATTTTATGAAGGGG 187628

RESULT 10
 AC127606
 LOCUS
 DEFINITION
 Rattus norvegicus clone CH230-20D15, *** SEQUENCING IN PROGRESS
 ***, 60 unordered pieces.
 AC127606
 VERSION
 AC127606.1 GI:21902814
 KEYWORDS
 HTG; HTG_PPHASEL.
 SOURCE
 ORGANISM
 Rattus norvegicus
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE AUTHORS

1 (bases 1 to 188670)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gablis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Hultik, S., Hune, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C.,
 Kratovic, J., Kucharski, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Louised, H.,
 Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhinney, E., McLeod, M.P., Meadow, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenko, S., Ogih, M., Okunolu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Prims, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojudo, K., Shen, H., Shooshnari, N., Sison, I.,
 Scherer, S., Scott, G., Shen, H., Shooshnari, N., Sison, I.,
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, K.,
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Unpublished
 2 (bases 1 to 188670)
 Direct Submission
 Morley, K.C.
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GXOD

QY	630	TCCTGTTTGACATTATATATTTCTTCCTTTACATTTGATTCATTAGCTAC	699		
Db 146805	TCCTGTTTGACATTATATTTCTTCCTTTACATTTGATTCATTAGCTAC	146864			
QY	690	TTTTTGACACCTTGATGAGCAATCAGACACGAAATTTACCACTGGAGTGTGTTT	749		
Db 146865	TTTTTGACACCTTGATGAGCAATCAGACACGAAATTTACCACTGGAGTGTGTTT	146924			
QY	750	GAGGAGAGATGTGATTTTATGTAAGGGGG	778		
Db 146925	GAGGAGAGATGTGATTTTATGTAAGGGGG	146953			
RESULT 11	AC096040	82586 bp	DNA	linear	HTG 11-JUL-2001
LOCUS	AC096040/C				
DEFINITION	Rattus norvegicus clone CH230-2911, *** SEQUENCING IN PROGRESS ***				
ACCESSION	AC096040				
VERSION	AC096040.4	GI:21723170			
KEYWORDS	HTG; HTGS; PHASE1.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Mumzy, D.M., Adams, C., Adio-Oduola, B., Ali-oman, F.R., Allen, C., Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayele, L., Banks, T., Barbarta, J., Benton, J., Blmage, K., Blankenburg, K., Bonini, D., Bouck, J., Bowle, S., Brieva, M., Brown, B., Brown, M., Bryant, N.P., Bubay, C., Burch, P., Burkett, C., Birtell, K.L., Byrd, N.C., Carron, I.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Dayla, M.K., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P., Gabriell, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, P., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E., Jacobson, B., Jay, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kralovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Lousseged, H., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McWhiney, E., McLeod, M.P., Medora, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, I., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Ogun, M., Okunolu, G., Oregunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulies, M., Ren, Y., Reyes, M., Rojas, A., Rojnbokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtrati, N., Sisson, I., Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerias, A., Tamerias, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umanai, K., Vazquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Woden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorritilla, S., Nelson, D., Weinstock, G., and Gibbs, R.				
TITLE	Direct Submision				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 82586)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-SEP-2001)				

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 82586)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17943701.

Center: Baylor College of Medicine
Contact address: DOW

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Center project name: GEGR

Center clone name: CH230-2911

Sequencing vector: Plasmid;

Assembly program: Phrap; version 0.990329

Consensus quality: 40324 bases at least Q30

consensus quality: 44113 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see <http://www.broad.com.tw/edu/decs/Genbank-3-4-14-11>)

(see http://www.ngsc.dcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This is a 'working draft' sequence. It currently consists of 46 contigs. The number of the next

* is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* This record will be undated with the finished comments

* as soon as it is available and the accession number will
* be preserved

* be preserved

```

*      1      1317: contig of 1317 bp in length
*      1318  1417: gap of unknown length

```

1418	contig of 1383 bp in length
2801	2900: gap of unknown length

```

* 2901 4244: contig of 1344 bp in length
* 4245 4344: gap of unknown length

```

```

*      4345      5864: contig of 1520 bp in length
*      5865      5964: gap of unknown length

```

```

* 5965 6968: contig of 1004 bp in length
* 6969 7068: gap of unknown length

```

**	7069	8166: contig of 1098 bp in length
**	8167	8266: gap of unknown length

*	8267	9384: contig of 1118 bp in length
*	9385	9484: gap of unknown length

*	9485	10630: contig of 1146 bp in length
*	10631	10730: gap of unknown length

10731	11858: contig of 1128 bp in length
11859	11958: gap of unknown length

*	11959	13413: contig of 1455 bp in length
*	13414	13513: gap of unknown length

*	13514	14847: contig of 1334 bp in length
*	14848	14947: gap of unknown length

*	14948	16180: contig of 1233 bp in length
*	16181	16280: gap of unknown length

16281	17804: contig of 1524 bp in length
17805	17904: gap of unknown length

17905	18913: contig of 1009 bp in length
18914	19013: gap of unknown length

19014	20050: contig of 1037 bp in length
20051	20150: gap of unknown length

20151	21378: contig of 1228 bp in length
21379	21478: gap of unknown length

21479	22828: contig of 1350 bp in length
22829	22928: gap of unknown length

22929	23998: contig of 1070 bp in length
23999	24098: gap of unknown length

24099	25596: contig of 1498 bp in length
25597	25696: gap of unknown length

```

* 25697 26965: contig of 1269 bp in length
* 26966 27065: gap of unknown length
* 27066 29291: contig of 2226 bp in length
* 29292 29391: gap of unknown length
* 29392 30870: contig of 1479 bp in length
* 30871 30970: gap of unknown length
* 30971 33590: contig of 2620 bp in length
* 33591 33690: gap of unknown length
* 33691 34829: contig of 1139 bp in length
* 34830 34929: gap of unknown length
* 34930 36600: contig of 1671 bp in length
* 36601 36700: gap of unknown length
* 36701 38157: contig of 1457 bp in length
* 38158 38257: gap of unknown length
* 38258 39358: contig of 1101 bp in length
* 39359 39458: gap of unknown length
* 39459 41104: contig of 1646 bp in length
* 41105 41204: gap of unknown length
* 41205 42418: contig of 1214 bp in length
* 42419 42518: gap of unknown length
* 42519 43862: contig of 1344 bp in length
* 43863 43962: gap of unknown length
* 43963 45961: contig of 1999 bp in length
* 45962 46061: gap of unknown length
* 46062 48791: contig of 2730 bp in length
* 48792 48891: gap of unknown length
* 48892 51571: contig of 2680 bp in length
* 51572 51671: gap of unknown length
* 51672 53160: contig of 1489 bp in length
* 53161 53260: gap of unknown length
* 53261 54713: contig of 1453 bp in length
* 54714 54813: gap of unknown length
* 54814 56686: contig of 1873 bp in length
* 56687 56786: gap of unknown length
* 56787 58525: contig of 1739 bp in length
* 58526 58625: gap of unknown length
* 58626 61362: contig of 2737 bp in length
* 61363 61462: gap of unknown length
* 61463 64676: contig of 3214 bp in length
* 64677 64776: gap of unknown length
* 64777 67299: contig of 2523 bp in length
* 67300 67399: gap of unknown length
* 67400 69326: contig of 1927 bp in length
* 69327 69426: gap of unknown length
* 69427 72187: contig of 2761 bp in length
* 72188 72287: gap of unknown length
* 72288 73531: contig of 1244 bp in length
* 73532 73631: gap of unknown length
* 73632 76752: contig of 3121 bp in length
* 76753 76852: gap of unknown length
* 76853 79280: contig of 2428 bp in length
* 79281 79380: gap of unknown length
* 79381 82586: contig of 3206 bp in length.

```

FEATURES

Source

1. 82586

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="CH230-2911"

BASE COUNT 23112 a 17332 c 15913 g 21644 t 4585 others

ORIGIN

Query Match 27.5%; Score 214.2; DB 2; Length 82586;

Best Local Similarity 92.2%; Pred. No. 1.2e-45;

Matches 248; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

OY 511 TTGACGTGTTTCCAGGAGAGAAAACATGGGTTGAATATAACAACTTCTGAACAT 570

Db 53049 TTGACGTGTTTCTCAATG-GAAAAACAAATGGGTTGAATATAACAACTTCTGAAT 52992

OY 571 TTATATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 629

Db 52991 TTATATATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 52992

```

OY 630 TCCTGTTGATGATTTATATTTCTCTTTTATCTGTTGATGATTTCTCACTTAGCTACA 689
Db 52931 TCCTGTTGATGATTTATATTTCTCTTTTATCTGTTGATGATTTCTCACTTAGCTACA 52872
OY 690 TTTTGGACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 749
Db 52871 TTTTGGACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 52812
OY 750 GAGAGAGATGATGATTTTATGAGAGGAG 778
Db 52811 GAGAGAGATGATGATTTTATGAGAGGAG 52783

```

RESULT 12

AL731811/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On Aug 13, 2002 this sequence version replaced gi:21912698.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: BM60A1

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 96% of reads

Chemistry: Dye-terminator Big Dye; 3% of reads

Consensus quality: 235503 bases at least Q40

Consensus quality: 236735 bases at least Q30

Consensus quality: 237530 bases at least Q20

Insert size: 238457; sum-of-contigs

Insert size: 241670; 2.5% error; agarose-fp

Quality coverage: 7.58x in Q20 bases; sum-of-contigs Quality coverage: 10.48x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 39133: contig of 39133 bp in length

39134 39233: gap of 100 bp

39234 84228: contig of 44995 bp in length

84229 84328: gap of 100 bp

84329 106119: contig of 21791 bp in length

106120 106219: gap of 100 bp

106220 119269: contig of 13050 bp in length

119270 119369: gap of 100 bp

119370 130432: contig of 11063 bp in length

130433 130532: gap of 100 bp

130533 218182: contig of 87650 bp in length

218183 218282: gap of 100 bp

218283 221609: contig of 3327 bp in length

221610 221709: gap of 100 bp

Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
9847074
2 (bases 1 to 36503)
Waterston, R.
The sequence of Homo sapiens cosmid clone U112B8
Unpublished (1999)
3 (bases 1 to 36503)
Waterston, R.
Direct Submission
Submitted (04-OCT-1996)
4 (bases 1 to 36503)
Waterston, R.
Direct Submission
Submitted (20-JAN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 36503)
Waterston, R.
Direct Submission
Submitted (27-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@waterston.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from one clone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was mapped by Grieff, M., Whyte, M. P., Thakker, R. V., and Mazzarelli, R. Sequence analysis of 139 kb in Xp22.1 containing spermine synthase and the 5' region of PRX. Genomics 44:227-231 (1997).

SOURCE INFORMATION:
This clone is from a chromosome X-specific cosmid library LLOXNCC01-U. The source of the chromosomes was a human/lameter hybrid, GM07397-F, from Robert Muesbaum at the University of Pennsylvania School of Medicine. Please contact the Lawrence Livermore National Laboratory at <http://www-bio.llnl.gov/genome> to obtain the clone.

VECTOR: Lawrie16
Location/Qualifiers
1. 36503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.1-22.2"
/clone="U112B8"
/clone_id="LLOXNCC01-U"
116. 196
/rpt_family="L2"
repeat_region
1980. 2046
/rpt_family="L2"
misc_feature
3690. 4103

misc_feature
3711. 3929
/note="match to EST AA211443 (NID:g1810130) zn55b01.g1"
/note="match to EST AA214031 (NID:g1812669) zn58f02.g1"
Complement (3980. 4103).
/note="match to EST AA211521 (NID:g1810175) zn55b01.r1"
misc_feature
repeat_region
4141. 4364
/rpt_family="L1"
4366. 4730
/rpt_family="MALR"
4732. 4953
/rpt_family="L1"
5596. 5711
/rpt_family="MALR"
6433. 6793
/rpt_family="L1"
7169. 7254
/rpt_family="L2"
7874. 8264
/rpt_family="L1"
8529. 8555
/rpt_family="L2"
10938. 11032
/rpt_family="MIR"
11095. 11460
/rpt_family="MALR"
11640. 11719
/rpt_family="L1"
12097. 12191
/rpt_family="L1"
12453. 12595
/rpt_family="L1"
13060. 13222
/rpt_family="MER1_type"
13286. 13385
/rpt_family="MER1_type"
13401. 13625
/rpt_family="MER1_type"
14107. 14274
/rpt_family="MIR"
14638. 14751
/rpt_family="L1"
14804. 15170
/rpt_family="MALR"
15277. 15396
/rpt_family="L2"
15512. 15625
/rpt_family="L2"
16941. 17238
/rpt_family="ALU"
17746. 18251
/rpt_family="MALR"
19251. 19321
/rpt_family="MER1_type"
19583. 19881
/rpt_family="ALU"
20137. 20192
/rpt_family="L1"
20198. 20256
/rpt_family="MER1_type"
20456. 20562
/rpt_family="MIR"
20820. 21010
/rpt_family="MER1_type"
23777. 23857
/rpt_family="MIR"
24315. 24471
/rpt_family="MER1_type"
24480. 24571
/rpt_family="MIR"
24736. 24842
/rpt_family="L2"
26372. 26608
/rpt_family="MIR"

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 19:13:41 ; Search time 77.739 Seconds
(without alignments)
8778.558 Million cell updates/sec

Title: US-09-647-019-1

Perfect score: 1 gctcctcagagctcagagagag.....ctgtgattttatcgaagggg 778

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391.6	50.3	824	10 US-09-880-192-4	Sequence 4, Appl1
2	226.2	29.1	466	10 US-09-960-352-5216	Sequence 5216, Ap
3	162.4	20.9	587	10 US-09-962-436-468	Sequence 468, Ap
4	41.2	5.3	4572	9 US-10-092-154-1971	Sequence 1971, Ap
5	41.2	5.3	4572	10 US-09-764-847-1971	Sequence 1971, Ap
6	37.2	4.8	147309	10 US-09-742-312-3	Sequence 3, Appl1
7	36.8	4.7	640681	10 US-09-790-988-1	Sequence 1, Appl1
8	36	4.6	1200	10 US-09-770-445-43	Sequence 43, Appl1
9	35.8	4.6	597	9 US-09-796-692-8553	Sequence 8553, Ap
10	35.8	4.6	24699	10 US-09-764-877-2419	Sequence 2419, Ap
11	34.6	4.4	1074	9 US-10-084-205-3	Sequence 3, Appl1
12	34.6	4.4	1074	10 US-09-925-637-3	Sequence 3, Appl1
13	34.6	4.4	1077	10 US-09-815-242-8282	Sequence 8282, Ap
14	34.6	4.4	7762	7 US-08-781-986A-124	Sequence 124, Ap
15	34.4	4.4	474	10 US-09-864-761-15062	Sequence 15062, A
16	34.2	4.4	1925	10 US-09-822-830A-340	Sequence 340, App
17	34.2	4.4	3592	10 US-09-777-745-1	Sequence 1, Appl1
18	34.2	4.4	3627	10 US-09-777-745-5	Sequence 5, Appl1
19	34	4.4	640681	10 US-09-790-988-1	Sequence 1, Appl1

20	33.8	4.3	1435	9 US-09-954-531-963	Sequence 963, App
21	33.6	4.3	368004	10 US-09-949-654-3	Sequence 3, Appl1
22	33.4	4.3	380	10 US-09-867-701-778	Sequence 778, App
23	33.4	4.3	1026	10 US-09-815-242-6560	Sequence 6560, Ap
24	33.4	4.3	6835	10 US-09-070-927A-204	Sequence 204, App
25	33.4	4.3	14176	10 US-09-764-864-1644	Sequence 1644, Ap
26	33.2	4.3	459	9 US-10-046-935-20	Sequence 20, Appl1
27	33.2	4.3	459	9 US-09-878-178-20	Sequence 20, Appl1
28	33.2	4.3	554	9 US-10-025-380-413	Sequence 413, App
29	33.2	4.3	554	10 US-09-922-217-413	Sequence 413, App
30	33.2	4.3	554	10 US-09-833-263-413	Sequence 413, App
31	33.2	4.3	1439	9 US-09-822-846-599	Sequence 599, App
32	33.2	4.3	1528	9 US-10-025-380-1106	Sequence 1106, App
33	33.2	4.3	1528	10 US-09-922-217-1106	Sequence 1106, App
34	33.2	4.3	1597	10 US-09-925-300-656	Sequence 656, App
35	33.2	4.3	2622	10 US-09-950-370-3	Sequence 3, Appl1
36	33.2	4.3	2706	10 US-09-950-370-13	Sequence 13, Appl1
37	33.2	4.3	2981	10 US-09-950-370-11	Sequence 11, Appl1
38	33.2	4.3	3065	10 US-09-960-352-10222	Sequence 10222, A
39	33	4.2	399	10 US-09-917-800A-1017	Sequence 1017, App
40	33	4.2	521	10 US-09-764-864-109	Sequence 109, App
41	33	4.2	302250	10 US-09-962-832-154	Sequence 154, App
42	33	4.2	323	10 US-09-983-965-1163	Sequence 2163, App
43	32.8	4.2	500	10 US-09-833-381-914	Sequence 914, App
44	32.8	4.2	500	10 US-10-098-841-330	Sequence 330, App
45	32.8	4.2	4205	9 US-10-098-841-330	Sequence 330, App

ALIGNMENTS

RESULT 1
US-09-880-192-4
Sequence 4, Application US/09880192
Patent No. US20020077470A1
GENERAL INFORMATION:
APPLICANT: Walker, Michael G.
APPLICANT: Volkmut, Wayne
APPLICANT: Klingler, Tod M.
APPLICANT: Azimzai, Yalda
TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
FILE REFERENCE: PB-0009-1 CIP
CURRENT FILING DATE: US/09/880,192
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PERL Program
SEQ ID NO 4
LENGTH: 824
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020077470A1 3601719CB1
US-09-880-192-4

QY	6	CAGACCTGAGAGAGACAGACAGCTCAGCTATTTCAGCCACATGAAAGACATGAAAT	65
DB	33	CAGTTCCTAATACCGGAGAGAGACAGAGCTATTTCAGCCATGAAAGACATGAAAT	92
QY	66	GAGATCCCGCTCAGAGAGACACCGGAGTTCCTCTAATCTGTAAAGCGCTTTTGTGT	125
DB	93	GAGATCCGAGCTCAGAGAGACACCGGAGTTCCTCTAATCTGTAAAGCGCTTTTGTGT	150
QY	126	TTTGACACTGCGCCCTCGGAGACTGCTCTCAGCAGTAAACCAATCCAGAGAGCGGCTA	185
DB	151	CTTGATCTGCTCCTCGGAGCTTCTTGGCAGTAAACCAATCAATTAAGAGGGATA	210
QY	186	AGACCTTGATATGTGAGAGACCAATTTCCAGTCCAGAGCCATCCAGGCGAATAT	245

Db 248 TTATCTTATTTTGGACCTCAGAGAACTATAGCCATTAATTGACACCTGGAGG 189
 QY 741 TGTGTTTGGAGAGATGATTTTATGA 772
 Db 188 TGTGTTTGGAGAGGATGATTTATGA 157

RESULT 4

US-10-092-154-1971/c
 ; Sequence 1971, Application US/10092154
 ; Publication No. US20030054375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC093C1
 ; CURRENT APPLICATION NUMBER: US/10/092.154
 ; CURRENT FILING DATE: 2002-03-07
 ; NUMBER OF SEQ ID NOS: 2003
 ; Prior Application removed - See File Wrapper or Palm
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1971
 ; LENGTH: 4672
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-092-154-1971

Query Match 5.3%; Score 41.2; DB 9; Length 4672;
 Best Local Similarity 49.1%; Pred. No. 0.15;
 Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 389 TTGTCACTGTCTGAGATCCAAATGTTAAAGTGAAGTGAATTTGCCCCAAAGGTG 448
 Db 3075 TTCTAAATAATCTCTACTGCTTCTATCATCAGGGGACAGGAATCTGGAAATAATAG 3016
 QY 449 AACACTAGTCGAAGACACAAAGTTCACATGATGATCTTGAATCAGAGATGATTT 508
 Db 3015 AGCGTAGAGTAAGGACATTTGAGATGATGATGATGATGATTTCTTTTAAATGATTT 2956
 QY 509 CGTTGACGTTTTCCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 568
 Db 2955 CTCAGAGTGATGATTTCTAATTTATCATATACATGATGATGATGATGATGATG 2896
 QY 569 ATTTATACATTTGTATGATGATCACAACCTCTGATGATGCC 610
 Db 2895 AATAAGTATTAATTTGTTCTATCATCAAAAACGATCTGAATGCC 2854

RESULT 5

US-09-764-847-1971/c
 ; Sequence 1971, Application US/09764847
 ; Patent No. US20020132767A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC009
 ; CURRENT APPLICATION NUMBER: US/09/764.847
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior Application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 2003
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1971
 ; LENGTH: 4672
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-847-1971

Query Match 5.3%; Score 41.2; DB 10; Length 4672;
 Best Local Similarity 49.1%; Pred. No. 0.15;
 Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 389 TTGTCACTGTCTGAGATCCAAATGTTAAAGTGAAGTGAATTTGCCCCAAAGGTG 448
 Db 3075 TTCTAAATAATCTCTACTGCTTCTATCATCAGGGGACAGGAATCTGGAAATAATGAG 3016

QY 449 AACAGTAGTCGAAGACACAAAGTTCACATGATGATGATGATGATGATGATGATG 508
 Db 3015 AGCGTAGAGTAAGGACATTTGAGATGATGATGATGATGATGATGATGATTTT 2956
 QY 509 CGTTGACGTTTTCCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 568
 Db 2955 CTCAGAGTGATGATTTCTAATTTATCATATACATGATGATGATGATGATGATG 2896
 QY 569 ATTTATACATTTGTATGATGATCACAACCTCTGATGATGCC 610
 Db 2895 AATAAGTATTAATTTGTTCTATCATCAAAAACGATCTGAATGCC 2854

RESULT 6

US-09-742-312-3/c
 ; Sequence 3, Application US/09742312
 ; Patent No. US20020045166A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHANDRAMOULISARAN, Iehwar et al
 ; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 ; TITLE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: CL000838
 ; CURRENT APPLICATION NUMBER: US/09/742.312
 ; CURRENT FILING DATE: 2000-12-22
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 147309
 ; TYPE: DNA
 ; ORGANISM: Human
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(147309)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-742-312-3

Query Match 4.8%; Score 37.2; DB 10; Length 147309;
 Best Local Similarity 50.0%; Pred. No. 19;
 Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 559 CTTCGTGACATTTTATACATTTGTATGATGATCACAACCTCTGATGATGCCAAGCTC 618
 Db 21129 CTGTTGAGATGACAGTACTAGATATGTTCAACACATCTTATCATCATCTGATTT 21070
 QY 619 TAGCAAAATATCTGTTTATCATTTATTTCTCTTACTGTTGATTTCTCA 678
 Db 21069 TAGAAAAAGCATATTAATTTATCTATTTTCTTAGTTTCTCAGATCTTACAGATA 21010
 QY 679 CTTAGCTACATTTTGGACCTTGTAGAGCAATGACACAGAAATTTACAACTGGGA 738
 Db 21009 CTGCTCTACAACTACAGAGCTTGTATCTCACCTCTCCAGAGTTCTTGACATATAGGG 20950
 QY 739 AGTGTG 744
 Db 20949 AGTATG 20944

RESULT 7

US-09-790-988-1/c
 ; Sequence 1, Application US/09790988
 ; Patent No. US20020127687A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHIGENOBU, SHUJI
 ; APPLICANT: WATANABE, HIDEMI
 ; APPLICANT: HATTORI, MASAHIRA
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
 ; FILE REFERENCE: 081356/0159
 ; CURRENT APPLICATION NUMBER: US/09/790.988
 ; CURRENT FILING DATE: 2001-02-23
 ; PRIOR APPLICATION NUMBER: JP2000-107160

Db 255 AAAATGATGATTTGTATACCCAGATTTTGGCTGATATTTGTATAGCATTTCTACAT 314
QY 593 ACAAACTCTGGAATGCCAAGACTCTAGCAAAATATCTGTTGTACATTTATATTC 652
Db 315 GTGAATTTATTAATACAAACTCCAGGCTAAACATCAATATTTCTTAATGCTT 374
QY 653 TTCTTTTACTGTTGGTCAATTTCTGACTTTAGCTAATTTTGGACCTTGAGA 707
Db 375 TATATTTTAAATGTTAAACCCCTATAGCCACTTTGGGAATGTTTAA 429

RESULT 10

US-09-764-877-2419/C
Sequence 2419, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2419
LENGTH: 24699
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-2419

Query Match 4.4%; Score 35.8; DB 10; Length 24699;
Best Local Similarity 50.3%; Pred. No. 17;
Matches 88; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 533 AAAAACAATGGTGAATAAACAATCTCGAACAATTTATATGATGATGATC 592
Db 12313 AAAATGATGATTTGTATACCCAGATTTTGGCTGATATTTGTATAGCATTTCTACAT 12254
QY 593 ACAAACTCTGGAATGCCAAGACTCTAGCAAAATATCTGTTGTACATTTATATTC 652
Db 12253 GTGAATTTATTAATACAAACTCCAGGCTAAACATCAATATTTCTTAATGCTT 12194
QY 653 TTCTTTTACTGTTGGTCAATTTCTGACTTTAGCTAATTTTGGACCTTGAGA 707
Db 12193 TATATTTTAAATGTTAAACCCCTATAGCCACTTTGGGAATGTTTAA 12139

RESULT 11

US-10-084-205-3/C
Sequence 3, Application US/10084205
Publication No. US20030049648A1
GENERAL INFORMATION:
APPLICANT: Choi, Gil
TITLE OF INVENTION: 37 *Staphylococcus aureus* Genes and Polypeptides
FILE REFERENCE: P515P1
CURRENT APPLICATION NUMBER: US/10/084,205
CURRENT FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: PCT/US00/23773
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: 60/151,933
PRIOR FILING DATE: 1999-09-01
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Patent Ver. 3.1
SEQ ID NO 3
LENGTH: 1074
TYPE: DNA
ORGANISM: *Staphylococcus aureus*
US-10-084-205-3

Query Match 4.4%; Score 34.6; DB 9; Length 1074;
Best Local Similarity 52.4%; Pred. No. 5.6;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 546 TTGAATTAACAACACTCTCGAACAATTTATATGATGATGATGATCAAACTCTGGA 605
Db 326 TTGTATCATATAGATCTTTAGAAATCAATATATTTTAAGCTCTTCTCAAGATTGGA 267
QY 606 ATGCCCAAGACTCTAGCAAAATATCTGTTGTACATTTATATTTCTCTTTACTTG 665
Db 266 AGTTCACCTTAATATCAATTAATCTCTCTTTTAACATTTCTACTCTCTTTATCA 207
QY 666 GTTGCAATTTCTCACTTATAGCTACAT 690
Db 206 GTCTCACTTAACATTTCTTCAATAT 182

RESULT 12

US-09-925-637-3/C
Sequence 3, Application US/09925637
Patent No. US2002010338A1
GENERAL INFORMATION:
APPLICANT: Choi
TITLE OF INVENTION: *Staphylococcus aureus* Polynucleotides and Polypeptides
FILE REFERENCE: P5560
CURRENT APPLICATION NUMBER: US/09/925,637
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/23773
PRIOR FILING DATE: 2000-08-31
PRIOR APPLICATION NUMBER: US 60/151,933
PRIOR FILING DATE: 1999-09-01
PRIOR APPLICATION NUMBER: US 08/781,986
PRIOR FILING DATE: 1997-01-03
PRIOR APPLICATION NUMBER: US 08/956,171
PRIOR FILING DATE: 1997-10-20
PRIOR APPLICATION NUMBER: US 60/009,861
PRIOR FILING DATE: 1996-01-06
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 3
LENGTH: 1074
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-637-3

Query Match 4.4%; Score 34.6; DB 10; Length 1074;
Best Local Similarity 52.4%; Pred. No. 5.6;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 546 TTGAATTAACAACACTCTCGAACAATTTATATGATGATGATGATCAAACTCTGGA 605
Db 326 TTGTATCATATAGATCTTTAGAAATCAATATATTTTAAGCTCTTCTCAAGATTGGA 267
QY 606 ATGCCCAAGACTCTAGCAAAATATCTGTTGTACATTTATATTTCTCTTTACTTG 665
Db 266 AGTTCACCTTAATATCAATTAATCTCTCTTTTAACATTTCTACTCTCTTTATCA 207
QY 666 GTTGCAATTTCTCACTTATAGCTACAT 690
Db 206 GTCTCACTTAACATTTCTTCAATAT 182

RESULT 13

US-09-815-242-8282/C
Sequence 8282, Application US/09815242
Patent No. US2002061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Onksen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes

FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8282
LENGTH: 1077
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1077)
US-09-815-242-8282

Query Match 4.4%; Score 34.6; DB 10; Length 1077;
Best Local Similarity 52.4%; Pred. No. 5.6;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 546 TTGAATTAACAACTCTCGTGAACATTTATACATTTGTATGATGATCAACAACTCTCTGA 605
DB 323 TTGTATCATTTAGATCTTTTGAATCAATATTTTAACTCTCTTCAAGATTGGA 264
QY 606 ATGCCAAGACTCTAGCAAAATATCCTGTTGTACATTATATTTCTCTTACTTG 665
DB 263 AGTTCAGCTTAATACCGCTACTTCCCTTTAAGATTCTACTTCTTATCATCA 204
QY 666 GTTGCATTTCTCACTTTAGTACAT 690
DB 203 GTTCACTTAACATTTCTTCAATAT 179

RESULT 14
US-08-781-986A-124
Sequence 124, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION: DATA
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 124:
SEQUENCE CHARACTERISTICS:
LENGTH: 7762 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-124

Query Match 4.4%; Score 34.6; DB 7; Length 7762;
Best Local Similarity 52.4%; Pred. No. 19;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 546 TTGAATTAACAACTCTCGTGAACATTTATACATTTGTATGATGATCAACAACTCTCTGA 605
DB 6730 TTGTATCATTTAGATCTTTTGAATCAATATTTTAACTCTCTTCAAGATTGGA 6789
QY 606 ATGCCAAGACTCTAGCAAAATATCCTGTTGTACATTATATTTCTCTTACTTG 665
DB 6790 AGTTCAGCTTAATACCGCTACTTCCCTTTAAGATTCTACTTCTTATCATCA 6849
QY 666 GTTGCATTTCTCACTTTAGTACAT 690
DB 6850 GTTCACTTAACATTTCTTCAATAT 6874

RESULT 15
US-09-864-761-15062
Sequence 15062, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

Wed Apr 16 05:42:52 2003

УВ-09-647-019-1.гггб

Page 7

```

1  PRIOR APPLICATION NUMBER: ECT/US01/00670
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: US 60/234,687
4  PRIOR FILING DATE: 2000-09-21
5  PRIOR APPLICATION NUMBER: US 09/608,408
6  PRIOR FILING DATE: 2000-06-30
7  PRIOR APPLICATION NUMBER: US 09/774,203
8  PRIOR FILING DATE: 2001-01-29
9  NUMBER OF SEQ. ID NOS: 49117
10 SOFTWARE: Annonamak Sequence Listing Engine vers. 1.1
11 SEQ ID NO 15062
12 LENGTH: 474
13 TYPE: DNA
14 ORGANISM: Homo sapiens
15 FEATURE:
16 OTHER INFORMATION: MAP TO AL022100.13
17 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
18 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.3
19 US-09-864-761-15062

```

Query Match 4.4%; Score 34.4; DB 10; Length 474;
 Best Local Similarity 46.0%; Pred. No. 3.9;
 Matches 16; Conservative 0; Mismatches 136; Indels 0; Gaps 0

Qy	400	TTGGAGATCCAAATGTTTAAAGTAACTGAAATTTGTCCCCAAAGGGAACAGTAATCG	45
Db	125	TTCAGATTGAAATTTTAAATCTATCAAGAGCAATATTTTCTTCCTAGAGCCTAGCA	18
Qy	460	AAAGACAACAAAGTTCACATGGATGCTTGTAACAGAGAAATGCATTTCGTGACGTG	515
Db	185	AAATGGAAGACCTTCACATATTTCTTGTATCACTGATGCTGATCTTCCAAAGCA	244
Qy	520	TTTTCCAGGGAGAAAAACAATGGGTTGAATTAACAATCTCTCGAACTATTATACAT	579
Db	245	ATTGACAAATGAGACATCTGAAGGCTTGATTTTGTATACGCTCAGTTTATATTTTT	304
Qy	580	TTGTATGATGATCAACAACTCTCGAATGCCAGACCTTGACAAAAATATCTGTGTG	639
Db	305	AGGACATCTTTCTCTTCACCTGGCTGAAAGCTCCCAAGAGCAAGGCTGTGTCTCTTCAT	364
Qy	640	ACATTATATTT	651
Db	365	CTTGATATTTCT	376

Search completed: April 16, 2003, 01:12:33
Job time : 344.739 secs

Page 7

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: April 9, 2003, 12:33:56 ; Search time 16.1089 Seconds
(without alignments)
155.253 Million cell updates/sec

Title: US-09-647-019-2

Perfect score: 440

Sequence: 1 MSKQPSINRAIOANINIM.....NLSEIQNKSELKFPVKGRQ 85

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum March 0%

Maximum March 100%

Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/PTUS-COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/backfill.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	17.3	941	4	US-07-757-022B-14
2	76	17.3	1022	4	US-07-757-022B-84
3	76	17.3	1038	4	US-07-757-022B-74
4	76	17.3	1049	4	US-07-757-022B-58
5	76	17.3	1140	4	US-07-757-022B-104
6	76	17.3	1270	4	US-07-757-022B-44
7	76	17.3	1311	4	US-07-757-022B-42
8	76	17.3	1313	4	US-07-757-022B-142
9	76	17.3	1314	4	US-07-757-022B-50
10	76	17.3	1320	4	US-07-757-022B-46
11	76	17.3	1354	4	US-07-757-022B-60
12	76	17.3	1354	4	US-07-757-022B-48
13	76	17.3	1361	4	US-07-757-022B-52
14	76	17.3	1363	4	US-07-757-022B-52
15	76	17.3	1404	4	US-07-757-022B-2
16	76	17.3	1404	4	US-07-757-022B-72
17	70.5	16.0	356	4	US-08-887-534A-72
18	67	15.2	328	1	US-08-414-926A-9
19	67	15.2	328	1	US-08-926-922-9
20	67	15.2	328	3	US-09-253-682-9
21	67	15.2	328	4	US-09-527-657-9
22	66.5	15.1	281	2	US-08-405-175A-9
23	66.5	15.1	392	3	US-08-301-162-2
24	66.5	15.1	392	4	US-09-461-240-2
25	66.5	15.1	392	4	US-09-968-927-2
26	66.5	15.1	428	3	US-08-301-162-18
27	66.5	15.1	428	4	US-09-461-240-18

28	66.5	15.1	428	4	US-09-968-927-18	Sequence 18, Appl
29	65.5	14.9	160	2	US-08-726-306A-35	Sequence 35, Appl
30	65.5	14.9	917	4	US-08-259-451-11	Sequence 11, Appl
31	64.5	14.7	667	4	US-09-303-064-55	Sequence 55, Appl
32	64.5	14.7	667	4	US-09-086-503-55	Sequence 55, Appl
33	64.5	14.7	1274	4	US-09-095-443-2	Sequence 2, Appl
34	63.5	14.4	708	4	US-09-413-814-18	Sequence 18, Appl
35	63.5	14.4	995	5	PCT-US95-04910-14	Sequence 14, Appl
36	63	14.3	244	1	US-08-696-827-1	Sequence 1, Appl
37	63	14.3	771	3	US-08-434-000A-8	Sequence 8, Appl
38	63	14.3	771	4	US-09-312-157-8	Sequence 8, Appl
39	63	14.3	892	4	US-08-857-076-42	Sequence 42, Appl
40	62.5	14.2	701	2	US-08-735-041A-6	Sequence 6, Appl
41	62.5	14.2	701	3	US-09-190-476B-6	Sequence 6, Appl
42	62.5	14.2	701	3	US-09-190-889A-6	Sequence 6, Appl
43	62.5	14.2	701	4	US-09-190-938B-6	Sequence 6, Appl
44	62	14.1	527	4	US-09-370-828-216	Sequence 216, Appl
45	62	14.1	549	3	US-08-886-886-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-07-757-022B-14
; Sequence 14, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gerner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Coe, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-14

Query Match 17.3%; Score 76; DB 4; Length 941;
Best Local Similarity 42.2%; Pred. No. 1.2;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPTEBGAFTTSEE-----KKPIPGMKKFPGP 63
DB 257 PKKPTTTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAP 301

RESULT 2
US-07-757-022B-84
Sequence 84, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-84

Query Match 17.3%; Score 76; DB 4; Length 1022;
Best Local Similarity 42.2%; Pred. No. 1.4;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
QY 29 PGAGOPRRKSTPTEBGAFTTSEE-----KKPIPGMKKFPGP 63

DB 338 PKKPTTTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAP 382

RESULT 3
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Gesner, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74

Query Match 17.3%; Score 76; DB 4; Length 1038;
Best Local Similarity 42.2%; Pred. No. 1.4;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
QY 25 PGAGOPRRKSTPTEBGAFTTSEE-----KKPIPGMKKFPGP 63
DB 322 PKKPTTTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAP 366

RESULT 4
US-07-757-022B-58
Sequence 58, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.


```

APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-5851
TELEFAX: (617)876-1170
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-58

Query Match 17.3%; Score 76; DB 4; Length 1049;
Best Local Similarity 42.2%; Pred. No. 1.4;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

OY 25 PGAGQPPRRKSTPTETGAPTSEE-----KKPIPGMKKPPGP 63
Db 365 PKEPTPTPKPAPPTTKPAPPTPKPAPPTPKPAPPTPKPAP 409

RESULT 5
US-07-757-022B-104
Sequence 104, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.

```

```

ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Caert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-104

Query Match 17.3%; Score 76; DB 4; Length 1140;
Best Local Similarity 42.2%; Pred. No. 1.6;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

OY 25 PGAGQPPRRKSTPTETGAPTSEE-----KKPIPGMKKPPGP 63
Db 456 PKEPTPTPKPAPPTTKPAPPTPKPAPPTPKPAPPTPKPAP 500

RESULT 6
US-07-757-022B-44
Sequence 44, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 1270 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-44

Query Match 17.3%; Score 76; DB 4; Length 1270;
Best Local Similarity 42.2%; Pred. No. 1.8;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

Qy 25 PGACQPPRRKSTPTEGAPTSEE-----KKPIGMMKKFPGP 63
Db 322 PKEPTPTTKPEAPATTKEPATTPEKPAPTAKPKKPAPTTPEKPA 366

RESULT 7
US-07-757-022B-42
Sequence 42, Application US/07757022B
Patent No. 6431142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
City: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-42

Query Match 17.3%; Score 76; DB 4; Length 1311;
Best Local Similarity 42.2%; Pred. No. 1.9;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

Qy 25 PGACQPPRRKSTPTEGAPTSEE-----KKPIGMMKKFPGP 63
Db 363 PKEPTPTTKPEAPATTKEPATTPEKPAPTAKPKKPAPTTPEKPA 407

RESULT 8
US-07-757-022B-142
Sequence 142, Application US/07757022B
Patent No. 6431142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
City: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 142:

SEQUENCE CHARACTERISTICS:
LENGTH: 1313 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-142

Query Match 17.3%; Score 76; DB 4; Length 1313;
Best Local Similarity 42.2%; Pred. No. 1.9;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGQPPRRKSTPTEGAPTTSE-----KKPIGMKKFGP 63
Db 365 PKERTPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAP 409

RESULT 9
US-07-757-022B-50
Sequence 50, Application US/07757022B
Patent No. 6431142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 17.3%; Score 76; DB 4; Length 1314;
Best Local Similarity 42.2%; Pred. No. 1.9;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGQPPRRKSTPTEGAPTTSE-----KKPIGMKKFGP 63
Db 366 PKERTPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAP 410

RESULT 10
US-07-757-022B-46
Sequence 46, Application US/07757022B
Patent No. 6431142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-46

Query Match 17.3%; Score 76; DB 4; Length 1320;
Best Local Similarity 42.2%; Pred. No. 1.9;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGQPPRRKSTPTEGAPTTSE-----KKPIGMKKFGP 63
Db 372 PKERTPTTKPEAPPTTKPEAPPTTKPEAPPTTKPEAP 416

RESULT 11
US-07-757-022B-60
Sequence 60, Application US/07757022B
Patent No. 6431142

APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-40

Query Match 17.3%; Score 76; DB 4; Length 1361;
Best Local Similarity 42.2%; Pred. No. 2;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

Qy 25 PGAGOPRRKSTPTETEGAPTSEE-----KKPIGKKKPPGP 63
Db 413 PKBPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAP 457

RESULT 14
US-07-757-022B-52
Sequence 52, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-52

Query Match 17.3%; Score 76; DB 4; Length 1363;
Best Local Similarity 42.2%; Pred. No. 2;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

Qy 25 PGAGOPRRKSTPTETEGAPTSEE-----KKPIGKKKPPGP 63
Db 415 PKBPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAP 459

RESULT 15
US-07-757-022B-2
Sequence 2, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-2

Query Match 17.3%; Score 76; DB 4; Length 1404;
Best Local Similarity 42.2%; Pred. No. 2.1;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

Oy 25 PGAGQPPRRKSTPRTBEGAPTTSE-----KKPIPGMKKFPGP 63
Db 456 PKPEPTTPKPEPAPTTKPEPAPTTKPEPAPPTAPKPKAPTTKPEPAP 500

Search completed: April 9, 2003, 12:40:29
Job time : 19.1089 secs

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 9, 2003, 12:31:10 ; Search time 11.6532 Seconds
(without alignments)
302.533 Million cell updates/sec

Title: US-09-647-019-2

Perfect score: 440
Sequence: 1 MSKQPSNVRAIQANINIPM.....NLSEIQVXSEKLVKGEQ 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	382.5	86.9	88	1 SMPX_HUMAN	Q9ubp9 homo sapien
2	78.5	17.8	1240	1 POLS_EBEV3	P27284 eastern equ
3	77.5	17.6	757	1 HT16_HYDAT	P33356 hydra atten
4	77	17.5	2142	1 BAT2_HUMAN	P48634 homo sapien
5	75.5	17.2	1239	1 POLS_EBEV	P08768 eastern equ
6	74.5	16.9	994	1 CLC1_MOUSE	Q64347 mus musculu
7	74.5	16.9	994	1 CLC1_RAT	P35524 ratu
8	73	16.6	1236	1 POLS_EBEV	P35524 ratu
9	72.5	16.5	1787	1 CHD3_CACEL	P23897 western equ
10	72	16.4	242	1 TONB_SALTY	Q22516 caenorhabdi
11	72	16.4	2517	1 NCR2_HUMAN	O9Y618 h nuclear r
12	71	16.1	381	1 TH11_MERJA	Q58341 methanococ
13	71	16.1	1509	1 GSRI_HUMAN	Q9nz44 homo sapien
14	70.5	16.0	1484	1 CES2_HUMAN	Q9bxc3 homo sapien
15	69.5	15.8	280	1 MACS_CHICK	P16527 gallus gall
16	68	15.5	239	1 TONB_ECOLI	P02929 escherichia
17	68	15.5	440	1 GAT4_RAT	P46152 ratu
18	68	15.5	558	1 CBX4_HUMAN	O00257 ratu
19	68	15.5	651	1 MEU6_SCHPO	Q94356 schizosach
20	67	15.2	1666	1 MYM1_YEAST	O62224 mus musculu
21	66.5	15.1	241	1 YG3D_YEAST	P53281 saccharomyc
22	66.5	15.1	982	1 POL_HTLV2	P03363 human t-cel
23	66.5	15.1	1152	1 MAP1_HUMAN	P27816 homo sapien
24	66	15.0	685	1 CSD_MYCPA	Q9K116 mycobacteri
25	65.5	14.9	774	1 NEK1_MOUSE	P51954 mus musculu
26	65.5	14.9	776	1 ANRS_HUMAN	Q9nu02 homo sapien
27	65.5	14.9	2365	1 CCAH_MOUSE	O88427 mus musculu
28	65.5	14.9	3038	1 TRIO_HUMAN	O75962 homo sapien
29	65	14.8	256	1 HAZB_MOUSE	P14434 mus musculu
30	65	14.8	416	1 NAPS_HUMAN	O14513 homo sapien
31	65	14.8	488	1 NDEB_ALCX	P94211 alca
32	65	14.8	537	1 MYPH_CHICK	O05623 gallus gall
33	65	14.8	732	1 TAU_MOUSE	P10637 mus musculu

34	65	14.8	751	1 TAU_RAT	P19332 ratu
35	64.5	14.7	914	1 BPRX_BACSU	P39793 bacillus su
36	64.5	14.7	1478	1 BCK1_YEAST	O01389 saccharomyc
37	64	14.5	229	1 HA23_MOUSE	P14439 mus musculu
38	64	14.5	1603	1 CALF_HUMAN	Q07092 homo sapien
39	64	14.5	3014	1 CLRI_HUMAN	Q9ny96 homo sapien
40	63.5	14.4	267	1 SSB2_HUMAN	P28290 homo sapien
41	63.5	14.4	679	1 RRP1_DROME	P27664 drosophila
42	63.5	14.4	2359	1 CCAH_RAT	O9660 ratu
43	63	14.3	128	1 YPRR_ECOLI	P03848 escherichia
44	63	14.3	255	1 HA21_MOUSE	P01904 mus musculu
45	63	14.3	255	1 HA22_MOUSE	P04224 mus musculu

ALIGNMENTS

RESULT 1
ID SMPX_HUMAN STANDARD; PRT; 88 AA.

AC Q9ubp9;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Small muscular protein (Stretch responsive skeletal muscle protein).
OS SMPX OR SRMX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065879; PubMed=10598820;
RA Patzack P., Zhuchenko O., Lee C.-C., Wehnert M.;
RT "Identification, mapping, and genomic structure of a novel X-
chromosomal human gene (SMPX) encoding a small muscular protein.";
Hum. Genet. 105:506-512(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Skeletal muscle;
RC Kemp T.V., Sadusky T.J., Carey N., Coulton G.R.;
RT "Identification of a stretch responsive skeletal muscle gene.";
Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RL [3]
RP SEQUENCE FROM N.A.
RA TISSUE=Skeletal muscle;
RC Strausberg R.;
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ database.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY AND ABUNDANTLY EXPRESSED IN
HEART AND SKELETAL MUSCLE.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL; AF129505; AAF19343.1; -;
CC EMBL; AJ250584; CAC08492.1; -;
CC EMBL; BC005948; AAH0548.1; -;
CC Genew; HGNC:11122; SMPX.
CC MIM; 300226; -;
CC SQ SEQUENCE 88 AA; 9559 MW; CE33D2839F0F9EB7 CRC64;

Query Match 86.9%; Score 382.5; DB 1; Length 88;

Best Local Similarity 86.0%; Pred. No. 9.3e-29; Index 1; Gaps 1;

Matches 74; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 1 MSKQPSNVRAIQANINIPGAFPGAGPGRKKESTPTEBGAFTTS-DEKRPIMKK 59

Db 3 MSKQPVNVAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPTSDSEKKPIPKAKK 62
 QY 60 PPGPVNLSEIQNVKSEIKFYPKGEQ 85
 Db 63 LPPGAVNLSEIQNVKSEIKFYPKAEQ 88

RESULT 2
 POLS_EEEV3 STANDARD; PRT; 1240 AA.
 ID POLS_EEEV3
 AC P27284;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polyprotein (P130) [contains: Coat protein C (EC 3.4.21.-)
 DE 6 kDa peptide; Spike glycoprotein E1].
 DE Eastern equine encephalitis virus (strain va33[ten broeck]) (Eastern
 OS equine encephalomyelitis virus).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OX NCBI_TaxID=11022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220727; PubMed=2024496;
 RA Weaver S.C., Scott T.W., Rico-Heese R.;
 RT "Molecular evolution of eastern equine encephalomyelitis virus in
 RT North America."
 RL Virology 182:774-784(1991).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M68094; AAA42980.1; --
 CC F1R; A39992; VHWVEV.
 CC HSP; P03315; IYCP.
 CC MEROPS; S03.001; --
 CC InterPro: IPR002548; Alpha_E1_glycop.
 CC InterPro: IPR000936; Alpha_E2_glycop.
 CC InterPro: IPR002533; Alpha_E3_glycop.
 CC InterPro: IPR001836; Alpha_core.
 CC InterPro: IPR000930; Togavirin.
 CC Pfam; PF00943; Alpha_E2_glycop; 1.
 CC Pfam; PF00944; Alpha_core; 1.
 CC Pfam; PF01563; Alpha_E3_glycop; 1.
 CC Pfam; PF01589; Alpha_E1_glycop; 1.
 CC PRINTS; PR00798; TOGAVIRIN.
 CC Coat protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
 KM Serine protease.
 FT CHAIN 1 260 COAT PROTEIN C.
 FT CHAIN 261 323 SPIKE GLYCOPROTEIN E3.
 FT CHAIN 324 743 SPIKE GLYCOPROTEIN E2.
 FT CHAIN 744 799 6 KDA PEPTIDE.
 FT CHAIN 800 1240 SPIKE GLYCOPROTEIN E1.
 FT ACT_SITE 137 137 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 211 211 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRNSMEM 259 276 POTENTIAL.
 FT TRNSMEM 695 712 POTENTIAL.
 FT TRNSMEM 722 736 POTENTIAL.
 FT TRNSMEM 781 799 POTENTIAL.
 FT TRNSMEM 1212 1236 POTENTIAL.

FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 933 933 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 1240 AA; 137290 MW; AEBEB159D083045 CRC64;

Query Match 17.8%; Score 78.5; DB 1; Length 1240;
 Best Local Similarity 33.3%; Pred. No. 9.1;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPTSDSEKKPIPKAKK 59
 Db 40 IEDLRRIANULTKORAPNPPAPRPAKRPAPKPKK---PAQAKKKRPPPAK 90

RESULT 3
 HT16 HYDAT STANDARD; PRT; 757 AA.
 ID HT16 HYDAT
 AC P53356;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase HTK16 (EC 2.7.1.112).
 GN HTK16
 OS Hydra attenuata (Hydra) (Hydra vulgaris).
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Anthomedusae;
 OC Hydrozoa; Hydra.
 OX NCBI_TaxID=6087;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Irvine;
 RX MEDLINE=94181280; PubMed=8134129;
 RA Chan T.A., Chu C.A., Rauen K.A., Krolier M., Tatarewicz S.M.,
 RA Steele R.E.;
 RT "Identification of a gene encoding a novel protein-tyrosine kinase
 RT containing SH2 domain and ankyrin-like repeats."
 RL Oncogene 9:1253-1259(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN SIGNAL TRANSDUCTION.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- TISSUE SPECIFICITY: EPITHELIAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00936; AAC27350.1; --
 CC HSP; P08631; 1AD5.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR000980; SH2.
 CC InterPro: IPR001245; Tyr_pkinase.
 CC Pfam; PF00017; SH2; 2.
 CC Pfam; PF00023; ank; 5.
 CC Pfam; PF00659; pkinase; 1.
 CC PRINTS; PR00109; TYRKINASE.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC ProDom; PD000093; SH2; 2.
 CC SMART; SM00248; ANK; 4.
 CC SMART; SM00252; SH2; 2.
 CC SMART; SM00219; TYKC; 1.
 CC PROSITE; PSS0088; ANK_REPEAT; 2.
 CC PROSITE; PSS0297; ANK_REP_REGION; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE, PSS0001; SH2. 2
 KW transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM SH2 domain; ANK repeat; Repeat.
 FT DOMAIN 10 102 SH2 1.
 FT REPEAT 115 147 ANK 1.
 FT REPEAT 151 180 ANK 2.
 FT REPEAT 184 214 ANK 3.
 FT REPEAT 219 248 ANK 4.
 FT REPEAT 252 281 ANK 5.
 FT REPEAT 287 379 SH2 2.
 FT DOMAIN 484 740 PROTEIN KINASE.
 FT NP BIND 490 498 ATP (BY SIMILARITY).
 FT BINDING 516 516 ATP (BY SIMILARITY).
 FT ACT SITE 608 608 BY SIMILARITY.
 FT MOD_RES 746 746 PHOSPHORYLATION (POTENTIAL).
 FT SEQUENCE 757 AA; 85598 MW; 14409391090794F CRC64;
 SO
 Query Match 17.6%; Score 77.5; DB 1; Length 757;
 Best Local Similarity 31.8%; Pred. No. 6.6; Mismatches 29; Indels 17; Gaps 4;
 Matches 27; Conservative 12;
 Oy 3 KQPSNVRAIQNINIPMGAFRPGAGOPRRKSTPTEEGAPT--TSEKKPIPGMKPF 60
 Db 399 KLPVPPSPRITKNNGLP-----QPLVPFTNESDSIFTRLCEKPKLP---KL 446
 Oy 61 PGVVNLSEIQNVKSELKFPKGEQ 85
 Db 447 PRPVNHTVEVPN---SVNVGQKGDQ 468
 RESULT 4
 BAT2 HUMAN STANDARD; PRT; 2142 AA.
 ID BAT2 HUMAN
 AC P48634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Large proline-rich protein BAT2 (HLA-B-associated transcript 2).
 GN BAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RX MEDLINE=90192810; PubMed=2156268;
 RA Banerji J., Sande J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 RT large proline-rich proteins with multiple repeated motifs and a
 RT single ubiquitin-like domain";
 RT Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378 (1990).
 RL [2]
 RN SEQUENCE OF 1-1860 FROM N.A.
 RX MEDLINE=9327029; PubMed=8499947; "Dense Alu clustering and a potential new member of the NF kappa B
 RA Tis P.J.M., Bouguetel L., Prieur S., Caterina D., Primas G.,
 RA Petre V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
 RA Cohen D.;
 RT "Dense Alu clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment";
 RL Nat. Genet. 3:137-145 (1993).
 CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC -1- FUNCTION: UNKNOWN.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.embnet.org/announcements>
 CC or send an email to license@ebi-emb.ch).
 CC EMBL, M33509; AAA35585.1; -

DR EMBL, M33518; AAA35586.1; JOINED.
 DR EMBL, M33512; AAA35586.1; JOINED.
 DR EMBL, Z15025; CAA78744.1; -.
 DR PIR, B35098; B35098.
 DR PIR, S36152; S36152.
 DR Genew, HGNC:13918; BAT2.
 DR MIM, 142580; -.
 DR Repeat.
 KM Repeat.
 FT DOMAIN 519 524 POLY-PRO.
 FT DOMAIN 636 657 GIN-RICH.
 FT DOMAIN 684 688 POLY-PRO.
 FT DOMAIN 699 704 POLY-PRO.
 FT DOMAIN 814 821 POLY-PRO.
 FT DOMAIN 1340 1345 POLY-GLY.
 FT DOMAIN 1398 1403 POLY-GLY.
 FT DOMAIN 1436 1442 POLY-PRO.
 FT DOMAIN 1982 1991 POLY-PRO.
 FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.
 FT REPEAT 41 95 1-1.
 FT REPEAT 98 154 1-1.
 FT REPEAT 281 337 1-2.
 FT REPEAT 1740 1795 1-4.
 FT REPEAT 337 549 2 X TYPE B REPEATS.
 FT REPEAT 337 549 2-1.
 FT REPEAT 476 549 2-2.
 FT REPEAT 1899 2089 3 X 50 AA TYPE C REPEATS.
 FT REPEAT 1899 1948 3-1.
 FT REPEAT 1965 2014 3-2.
 FT REPEAT 2040 2089 3-3.
 FT REPEAT 57 57 R -> A (IN REF. 2).
 FT CONFLICT 109 109 Q -> S (IN REF. 2).
 FT CONFLICT 414 414 P -> P (IN REF. 2).
 FT CONFLICT 532 532 T -> K (IN REF. 2).
 FT CONFLICT 682 682 Q -> K (IN REF. 2).
 FT CONFLICT 730 730 E -> D (IN REF. 2).
 FT CONFLICT 750 750 L -> R (IN REF. 2).
 FT CONFLICT 834 834 A -> T (IN REF. 2).
 FT CONFLICT 1035 1035 G -> A (IN REF. 2).
 FT CONFLICT 1068 1068 M -> L (IN REF. 2).
 FT CONFLICT 1285 1285 P -> R (IN REF. 2).
 FT CONFLICT 1400 1400 G -> A (IN REF. 2).
 FT CONFLICT 1611 1611 T -> S (IN REF. 2).
 FT CONFLICT 1729 1729 G -> A (IN REF. 2).
 SO SEQUENCE 2142 AA; 227840 MW; 32DDF16B9B52420A CRC64;
 Query Match 17.5%; Score 77; DB 1; Length 2142;
 Best Local Similarity 30.4%; Pred. No. 23; Mismatches 24; Conservative 10; Indels 14; Gaps 3;
 Oy 3 KQPSNVRAIQNINIPMGAFRPGAGOPRRKSTPTEEGAPTSEKKPIPGMKPF 55
 Db 875 KEETAQLTGPGAGKLP--ASRGAGPPPPRRSRRTETMG-PPGSSNRGIPPEEGAP 931
 Oy 56 ----GMKKFPGVVNLSEI 70
 Db 932 PRRAGPIKKPPPKVEEL 950
 RESULT 5
 ID POLS EREV STANDARD; PRT; 1239 AA.
 AC P08768;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [containing: Coat protein C (EC 3.4.21.-)
 DE (Capsid protein C), Spike glycoprotein E3, Spike glycoprotein E2,
 DE 6 kDa peptide, Spike glycoprotein E1].
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
 OS virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus
 OX NCBI_TaxID=11021;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=82V-2137;
RX MEDLINE=87282265; PubMed=2886548;
RA Chang C.-J.J., Trent D.W.;
RT "Nucleotide sequence of the genome region encoding the 26S mRNA of
RT eastern equine encephalomyelitis virus and the deduced amino acid
RL sequence of the viral structural proteins.";
J. Gen. Virol. 68:2129-2142(1987).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDA POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HENAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05816; CAA29261.1; -
DR PIR: A26816; VMVEE.
DR HSSP: P03315; IVCP.
DR MEROPS: S03.001; -.
DR InterPro: IPR002548; Alpha E1 glycop.
DR InterPro: IPR000936; Alpha E2 glycop.
DR InterPro: IPR002533; Alpha E3 glycop.
DR InterPro: IPR001836; Alpha core.
DR InterPro: IPR000930; Togavirin.
DR Pfam: PF00943; Alpha_E2_glycop.1.
DR Pfam: PF00944; Alpha_core.1.
DR Pfam: PF01563; Alpha_E3_glycop.1.
DR Pfam: PF01589; Alpha_E1_glycop.1.
DR PRINTS: PR00798; TOGAVIRIN.
DR Coats protein; Polypeptide; Transmembrane; glycoprotein; Hydrolase;
KW Serine protease.
FT CHAIN 1 259 COAT PROTEIN C.
FT CHAIN 260 322 SPIKE GLYCOPROTEIN E3.
FT CHAIN 323 742 SPIKE GLYCOPROTEIN E2.
FT CHAIN 743 798 6 KDA PEPTIDE.
FT CHAIN 799 1239 SPIKE GLYCOPROTEIN E1.
FT ACT_SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT TRANSMEM 261 277 POTENTIAL.
FT TRANSMEM 684 701 POTENTIAL.
FT TRANSMEM 727 737 POTENTIAL.
FT TRANSMEM 777 798 POTENTIAL.
FT TRANSMEM 1211 1235 POTENTIAL.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 932 932 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1239 AA; 137431 MW; 8C7664A405D241C CRC64;

Query Match 17.2%; Score 75.5; DB 1; Length 1239;
Best Local Similarity 31.1%; Pred. No. 17;
Matches 19; Conservative 13; Mismatches 22; Indels 7; Gaps 2;

```

```

ID CLC1 MOUSE STANDARD; PRT; 994 AA.
AC 064367;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chloride channel protein, skeletal muscle (Chloride channel protein
DE 1) (Clc-1).
GN CLCN1 OR CLC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=97468267; PubMed=9321463;
RA Schnuelle V., Antropova O., Gronemeyer M., Wedemeyer N., Jockusch H.,
RA Bartsch J.W.;
RT "The mouse Clc1/myotonia gene: Ethn insertion, a variable AATC repeat,
RT and PCR diagnosis of alleles.";
RL Mamm. Genome 8:718-725(1997).
RP [2]
RP SEQUENCE OF 350-467 FROM N.A.
RX MEDLINE=92065955; PubMed=165965;
RA Steinmeyer K., Klocke R., Ortland C., Gronemeyer M., Jockusch H.,
RA Gruender S., Jentsch T.J.;
RT "Inactivation of muscle chloride channel by transposon insertion in
RT myotonic mice.";
RL Nature 354:304-308(1991).
CC -1- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE CHLORIDE CHANNEL FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z95127; CAB08359.1; -
DR EMBL: Z95128; CAB08359.1; JOINED.
DR EMBL: Z95129; CAB08359.1; JOINED.
DR EMBL: Z95130; CAB08359.1; JOINED.
DR EMBL: Z95131; CAB08359.1; JOINED.
DR EMBL: Z95132; CAB08359.1; JOINED.
DR EMBL: Z95133; CAB08359.1; JOINED.
DR EMBL: Z95134; CAB08359.1; JOINED.
DR EMBL: Z95135; CAB08359.1; JOINED.
DR EMBL: Z95136; CAB08359.1; JOINED.
DR EMBL: Z95137; CAB08359.1; JOINED.
DR EMBL: Z95138; CAB08359.1; JOINED.
DR EMBL: Z95139; CAB08359.1; JOINED.
DR EMBL: Z95140; CAB08359.1; JOINED.
DR EMBL: Z95141; CAB08359.1; JOINED.
DR EMBL: Z95142; CAB08359.1; JOINED.
DR EMBL: Z95143; CAB08359.1; JOINED.
DR EMBL: Z95144; CAB08359.1; JOINED.
DR EMBL: Z95145; CAB08359.1; JOINED.
DR EMBL: Z95146; CAB08359.1; JOINED.
DR EMBL: Z95147; CAB08359.1; JOINED.
DR EMBL: Z95148; CAB08359.1; JOINED.
DR EMBL: Z95149; CAB08359.1; JOINED.
DR EMBL: X62897; CAA44686.1; -
DR EMBL: X62898; CAA44686.1; -
DR EMBL: X62899; CAA44686.1; -
DR EMBL: X62896; CAA44685.1; -
DR WGI: 88417; Clcn1.

```

RESULT 6%
CLC1_MOUSE

```

DR InterPro: IPR000644; CBS domain.
DR InterPro: IPR001807; Cl-channel_volt.
DR Pfam: PF00571; CBS; 2.
DR Pfam: PF00654; Voltage CLC; 1.
DR PRINTS: PR00762; CLCHANNEL.
DR SMART: SM00116; CBS; 1.
KM Ionic channel; Ion transport; Voltage-gated channel; Transmembrane;
KW CBS domain; Repeat.
FT DOMAIN 1 117 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 118 137 1 (POTENTIAL).
FT TRANSMEM 159 182 2 (POTENTIAL).
FT TRANSMEM 206 227 3 (POTENTIAL).
FT TRANSMEM 235 254 4 (POTENTIAL).
FT TRANSMEM 266 290 5 (POTENTIAL).
FT TRANSMEM 305 323 6 (POTENTIAL).
FT TRANSMEM 348 368 7 (POTENTIAL).
FT TRANSMEM 391 414 8 (POTENTIAL).
FT TRANSMEM 456 475 9 (POTENTIAL).
FT TRANSMEM 478 496 10 (POTENTIAL).
FT TRANSMEM 524 545 11 (POTENTIAL).
FT TRANSMEM 553 572 12 (POTENTIAL).
FT DOMAIN 573 844 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 845 863 13 (POTENTIAL).
FT DOMAIN 864 994 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 607 662 CBS 1.
FT DOMAIN 825 877 CBS 2.
SQ SEQUENCE 994 AA; 109993 MW; 2136CB413D58D CRC64;

Query Match 16.9%; Score 74.5; DB 1; Length 994;
Best Local Similarity 34.0%; Pred. No. 17;
Matches 18; Conservative 7; Mismatches 17; Indels 11; Gaps 2;

QY 19 PMGAFR-----PGAGPPRRKSTPTETEGAPTSEKKPIGMKKFPPG 63
DB 889 PLASFRNTTSIRKTPGPPPPASWNVPEGEDGAP---EREVWPTMPTVP 938

RESULT 7
CLC1_RAT STANDARD; PRT; 994 AA.
ID CLC1_RAT
AC P35524;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chloride channel protein, skeletal muscle (Chloride channel protein
DE 1) (CLC-1).
GN CLCN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=92065954; PubMed=1659664;
RA Steinmeyer K., Ottland C., Jentsch T.J.;
RT "Primary structure and functional expression of a developmentally
RT regulated skeletal muscle chloride channel.";
RL Nature 354:301-304 (1991).
CC -1- FUNCTION: VOLTAGE-GATED CHLORIDE CHANNEL. CHLORIDE CHANNELS HAVE
CC SEVERAL FUNCTIONS INCLUDING THE REGULATION OF CELL VOLUME;
CC MEMBRANE POTENTIAL STABILIZATION, SIGNAL TRANSDUCTION AND
CC TRANSEPITHELIAL TRANSPORT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN SKELETAL MUSCLES.
CC -1- SIMILARITY: BELONGS TO THE CHLORIDE CHANNEL FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62894; CA44683.1; -.
DR PIR: S19595; S19595.
DR InterPro: IPR000644; CBS domain.
DR InterPro: IPR001807; Cl-channel_volt.
DR Pfam: PF00571; CBS; 2.
DR Pfam: PF00654; Voltage CLC; 1.
DR PRINTS: PR00762; CLCHANNEL.
KM Ionic channel; Ion transport; Voltage-gated channel; Transmembrane;
KW CBS domain; Repeat.
FT DOMAIN 1 117 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 118 137 1 (POTENTIAL).
FT TRANSMEM 159 182 2 (POTENTIAL).
FT TRANSMEM 206 227 3 (POTENTIAL).
FT TRANSMEM 235 254 4 (POTENTIAL).
FT TRANSMEM 266 290 5 (POTENTIAL).
FT TRANSMEM 305 323 6 (POTENTIAL).
FT TRANSMEM 348 368 7 (POTENTIAL).
FT TRANSMEM 391 414 8 (POTENTIAL).
FT TRANSMEM 456 475 9 (POTENTIAL).
FT TRANSMEM 478 496 10 (POTENTIAL).
FT TRANSMEM 524 545 11 (POTENTIAL).
FT TRANSMEM 553 572 12 (POTENTIAL).
FT DOMAIN 573 844 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 845 863 13 (POTENTIAL).
FT DOMAIN 864 994 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 607 662 CBS 1.
FT DOMAIN 825 877 CBS 2.
SQ SEQUENCE 994 AA; 110073 MW; DCDDCD026E48FAE CRC64;

Query Match 16.9%; Score 74.5; DB 1; Length 994;
Best Local Similarity 34.0%; Pred. No. 17;
Matches 18; Conservative 7; Mismatches 17; Indels 11; Gaps 2;

QY 19 PMGAFR-----PGAGPPRRKSTPTETEGAPTSEKKPIGMKKFPPG 63
DB 889 PLASFRNTTSIRKTPGPPPPASWNVPEGEDGAP---EREVWPTMPTVP 938

RESULT 8
POLS_WEEV STANDARD; PRT; 1236 AA.
ID POLS_WEEV
AC P13897; Q88696; Q88697; Q88698; Q88699; Q88700;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Structural polyprotein (P130) [Contains Coat protein C (BC 3.4.21.-)
DE (Capaid protein C); Spike glycoprotein E1].
DE 6 kDa peptide; Spike glycoprotein E1].
OS Western equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX NCBI_TaxID=11039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BFS1703;
RX MEDLINE=88320369; PubMed=3413072;
RA Hahn C.S., Luestig S., Straus E.G., Straus J.H.;
RT "Western equine encephalitis virus is a recombinant virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5997-6001 (1988).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL, J03854; AAA42999.1; -.
 DR PIR, A35587; VHMVME.
 DR HSSP, P03315; 1VCP.
 DR MEROPS, S03.001; -.
 DR InterPro, IPR002548; Alpha_E1_glycop.
 DR InterPro, IPR000936; Alpha_E2_glycop.
 DR InterPro, IPR002533; Alpha_E3_glycop.
 DR InterPro, IPR001836; Alpha_E3_glycop.
 DR InterPro, IPR000930; Togavirin.
 DR Pfam, PF00943; Alpha_E2_glycop; 1.
 DR Pfam, PF00944; Alpha_E2_glycop; 1.
 DR Pfam, PF01563; Alpha_E3_glycop; 1.
 DR Pfam, PF01589; Alpha_E1_glycop; 1.
 DR PRINTS, PR00796; TOGAVIRIN.
 DR Coar protein; Polyprotein; Transmembrane; Glycoprotein; Hydrolase;
 KW Serine protease.
 FT CHAIN 1 259 COAT PROTEIN C.
 FT CHAIN 260 319 SPIKE GLYCOPROTEIN E3.
 FT CHAIN 320 742 SPIKE GLYCOPROTEIN E2.
 FT CHAIN 743 797 6 KDA PEPTIDE.
 FT CHAIN 798 1236 SPIKE GLYCOPROTEIN E1.
 FT ACT_SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRANSMEM 684 702 POTENTIAL.
 FT TRANSMEM 719 737 POTENTIAL.
 FT TRANSMEM 758 775 POTENTIAL.
 FT TRANSMEM 777 793 POTENTIAL.
 FT TRANSMEM 1206 1227 POTENTIAL.
 FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 936 936 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1042 1042 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1236 AA; 136082 MW; 0756DB0B0A1CCA96 CRC64;
 Query Match 16.6%; Score 73; DB 1; Length 1236;
 Best Local Similarity 26.7%; Pred. No. 29;
 Matches 16; Conservative 14; Mismatches 24; Indels 6; Gaps 1;
 QY 6 ISNVAIQTANINIPGAFRPGAGOPRRKREKSTPETEBGAPTTSEB-----KKPIPGKK 59
 DB 41 IEDLRRIANILFKQSRSPNPPEPGPKKKKAPKPKTPQPKKKKQQAKTTRKPKPKRQ 100
 RESULT 9
 CHD3_CABEL STANDARD; PRT; 1787 AA.
 AC Q22516; Q18794;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain helicase-DNA-binding protein 3 homolog (CHD-3).
 OS CHD-3 OR T1468.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_Taxid=6239;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=20530482; PubMed=11076750;
 RA von Zellewsky T., Palladino F., Brunschwig K., Tobler H., Hajnal A.,
 RA Mueller F.;
 RT "The C. elegans Mi-2 chromatin-remodelling proteins function in vulval
 RT cell fate determination.";
 RL Development 127:5277-5284(2000).
 RC SEQUENCE FROM N.A.
 RC STRAIN=Br1sc1 N2;

RA Matthews P., McMurtry A.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: Chromatin-remodelling protein that function in vulval
 CC cell fate determination.
 CC - SUBCELLULAR LOCATION: Nuclear (potential).
 CC - SIMILARITY: BELONGS TO THE SNF2/RAO54 HELICASE FAMILY.
 CC - SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC - SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC -----
 DR EMBL, AF308444; AAG29837.1; -.
 DR EMBL, Z67884; CAA91810.1; -.
 DR EMBL, Z67881; CAA91810.1; JOINED.
 DR EMBL, Z67881; CAA91798.1; -.
 DR EMBL, Z67884; CAA91798.1; JOINED.
 DR WormPep, T14G8.1; CE03657.
 DR InterPro, IPR000953; Chromo.
 DR InterPro, IPR004410; DEAD.
 DR InterPro, IPR004464; DEAD box.
 DR InterPro, IPR001650; Helicase_C.
 DR InterPro, IPR000330; SNF2_N.
 DR InterPro, IPR001965; Znf PHD.
 DR InterPro, IPR001841; Znf PHD.
 DR Pfam, PF00176; SNF2_N.1.
 DR Pfam, PF00271; Helicase_C.1.
 DR Pfam, PF00385; Chromo.1.
 DR Pfam, PF00628; PHD.2.
 DR SMART, SM00298; CHROMO.2.
 DR SMART, SM00487; DEXDC; 1.
 DR SMART, SM00490; HELIC; 1.
 DR SMART, SM00249; PHD.2.
 DR SMART, SM00184; RING.2.
 DR PROSITE, PS00598; CHROMO.1; FALSE_NEG.
 DR PROSITE, PS50013; CHROMO.2; 1.
 DR PROSITE, PS00690; DEAD ATP HELICASE; 1.
 DR PROSITE, PS01359; ZF PHD.1; 2.
 DR PROSITE, PS0016; ZF PHD.2; 2.
 DR Chromatin regulator; Nuclear protein; Repeat; Helicase; DNA-binding;
 KW ATP-binding; Zinc-finger.
 FT DOMAIN 59 62 POLY-LYS.
 FT ZN_FING 265 312 PHD-TYPE 1.
 FT ZN_FING 328 375 PHD-TYPE 2.
 FT DOMAIN 373 476 CHROMO.1.
 FT DOMAIN 501 583 CHROMO.2.
 FT DOMAIN 1287 1291 POLY-ARG.
 FT NP_BIND 641 648 ATP (POTENTIAL).
 FT SITE 763 766 DEAD BOX.
 SQ SEQUENCE 1787 AA; 205254 MW; 1EFC1EFCES9740 CRC64;
 Query Match 16.5%; Score 72.5; DB 1; Length 1787;
 Best Local Similarity 40.8%; Pred. No. 48;
 Matches 20; Conservative 3; Mismatches 11; Indels 15; Gaps 2;
 QY 35 ESTPETER--GAPTTSEKKPIPGKK-----PPGPVINS 68
 DB 36 EEEVTEESQGVPTTSEKKKPPKKGKSSKKKNNCDYDPYKST 84
 RESULT 10
 TONB_SALTY STANDARD; PRT; 242 AA.
 AC P25945;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE TonB protein.

GN TONB OR STM1737.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OK NCBI_TaxID=602;
 RN (1)
 RX SEQUENCE FROM N.A.
 RA MEDLINE=91094049; PubMed=2266561;
 RA Hannavy K., Barr G.C., Dorman C.J., Adamson J., Mazengera L.R.,
 RA Gallagher M.P., Evans J.S., Levine B.A., Trayer I.P., Higgins C.F.;
 RT "Tomb protein of Salmonella typhimurium. A model for signal
 RT transduction between membranes.";
 RL J. Mol. Biol. 216:897-910(1990).
 RN (2)
 RP REVISIONS TO 42: 58-60 AND 168.
 RX MEDLINE=93302513; PubMed=8316087;
 RA Karlsson M., Hannavy K., Higgins C.F.;
 RT "A sequence-specific function for the N-terminal signal-like sequence
 RT of the tomb protein.";
 RL Mol. Microbiol. 8:379-386(1993).
 RN (3)
 RN SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SSGC112 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
 CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
 CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES SUCH AS COBALAMIN,
 CC AND VARIOUS IRON COMPOUNDS (SUCH AS IRON DICITRATE, ENTEROCHELIN,
 CC AERODACTIN, ETC.). IN THE ABSENCE OF TONB THESE RECEPTORS BIND
 CC THEIR SUBSTRATES BUT DO NOT CARRY OUT ACTIVE TRANSPORT. TONB ALSO
 CC INTERACTS WITH SOME COLICINS AND IS INVOLVED IN THE ENERGY-
 CC DEPENDENT, IRREVERSIBLE STEPS OF BACTERIOPHAGES PH1-80 AND T1
 CC INFECTION. IT COULD ACT TO TRANSDUCE ENERGY FROM THE CYTOPLASMIC
 CC MEMBRANE TO SPECIFIC ENERGY-REQUIRING PROCESSES IN THE OUTER
 CC MEMBRANE, RESULTING IN THE RELEASE INTO THE PERIPLASM OF LIGANDS
 CC BOUND BY THESE OUTER MEMBRANE PROTEINS.
 CC -1- SUBUNIT: HOMODIMER. FORMS A COMPLEX WITH THE ACCESSORY PROTEINS
 CC EXBA AND EXBD (By similarity)
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
 CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
 CC PERIPLASM.
 CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1ab-sib.ch/announce/>
 CC or send an email to license@1ab-sib.ch).
 CC -----
 DR EMBL, X56434; CAJ39818.1; -
 DR EMBL, AE008777; AL20655.1; -
 DR PIR, S13257; S13257.
 DR HSSP, P94739; 1IHR.
 DR SEVGENE, SG10391; TONB.
 DR InterPro, IPR00538; TONB.
 DR Pfam, PF03544; TONB; 1.
 DR PRINTS, PR01374; TONBPROTEIN.
 KM TRANSPORT; Protein transport; Bacteriocin transport; Inner membrane;
 KM Periplasmic; Transmembrane; Signal-anchor; Repeat; Phage recognition;
 KM Complete proteome.
 FT TRANSMEM 1 32 SIGNAL-ANCHOR (BY SIMILARITY).
 FT DOMAIN 33 242 PERIPLASMIC (BY SIMILARITY).
 FT DOMAIN 70 83 7 X 2 AA TANDEM REPEATS OF E-P.

FT DOMAIN 93 106 7 X 2 AA TANDEM REPEATS OF K-P.
 SQ SEQUENCE 242 AA; 26293 MW; 7648F6BB3150FA6B CRC64;
 Query Match 16.4%; Score 72; DB 1; Length 242;
 Best Local Similarity 30.4%; Pred. No. 6.2;
 Matches 24; Conservative 8; Mismatches 41; Indels 6; Gaps 2;
 QY 4 QPISNVAIQANINIMGAFRP3---ACQPPRKSTPTEEGAPPTSEKKPIGMKK 59
 DB 43 QPITVMSPADLEPPQAVQPPPEVPEBPPEPPPIPEPPKAVVLEKPKPKPKP 102
 QY 60 FPGPVNLSIQNVKSELK 78
 DB 103 KPKPVKKVEE--QPKREV 119
 RESULT 11
 ID NCR2 HUMAN STANDARD; PRT; 2517 AA.
 AC Q9Y618; Q9Y5U0; Q13354; O00613; O15416;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of
 DE retinoic acid and thyroid hormone receptor) (SMRT) (SMRTE) (Thyroid-
 DE retinoic-acid-receptor-associated co-repressor) (T3 receptor-
 DE associating factor) (TRAC) (CTG26).
 GN NCOR2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CC NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUE=Pluticary;
 RX MEDLINE=99178941; PubMed=10077563;
 RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM SMRT).
 RC TISSUE=Cervical adenocarcinoma;
 RX MEDLINE=99199215; PubMed=10097068;
 RA Park E.J., Schreien D.J., Yang M., Li H., Li L., Chen J.D.;
 RT "SMRTE, a silencing mediator for retinoid and thyroid hormone
 RT receptors-extended isoform that is more related to the nuclear
 RT receptor corepressor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
 RN (3)
 RP SEQUENCE OF 1023-2517 FROM N.A.
 RC TISSUE=Cervical adenocarcinoma;
 RX MEDLINE=96008552; PubMed=7566127;
 RA Chen J.D., Evans R.W.;
 RT "A transcriptional co-repressor that interacts with nuclear hormone
 RT receptors.";
 RL Nature 377:454-457(1995).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM TRAC-1).
 RC TISSUE=Fetal liver;
 RX MEDLINE=96408715; PubMed=8813722;
 RA Sande S., Privalsky M.L.;
 RT "Identification of TRACs (T3 receptor-associated cofactors), a family
 RT of cofactors that associate with, and modulate the activity of,
 RT nuclear hormone receptors.";
 RL Mol. Endocrinol. 10:813-825(1996).
 RN (5)
 RP SEQUENCE OF 428-613 FROM N.A.
 RC TISSUE=Brain cortex;
 RX MEDLINE=97369492; PubMed=9225980;
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kildat A.S.,
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
 RT "CDNAs with long CAG trinucleotide repeats from human brain.";
 RL Hum. Genet. 100:114-122(1997).

CC -1- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION
CC -1- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH FTIIB.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SMRT/TRAC-2 (SHOWN HERE) AND
CC TRAC-1; ARE PRODUCED BY ALTERNATIVE SPLICING. TRAC-1 CONTAINS ONLY
CC THE C-TERMINAL RECEPTOR-INTERACTING DOMAIN AND ACTS AS AN
CC ANTI-REPRESSOR.
CC -1- TISSUE SPECIFICITY: Ubiquitous. HIGH LEVELS OF EXPRESSION ARE
CC DETECTED IN LUNG, SPLEEN AND BRAIN.
CC -1- INDUCTION: REGULATED DURING CELL CYCLE PROGRESSION.
CC -1- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC DOMAINS (ID1 AND ID2).
CC -1- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -1- SIMILARITY: CONTAINS 1 SANT-A DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CORN BOXES.
CC -1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF113003; AAD20946.1; -
CC EMBL; AF125672; AAD22973.1; -
CC EMBL; U37146; AAC50236.1; -
CC EMBL; S83390; AAB50847.1; -
CC EMBL; U80750; AAB91446.1; -
CC TRANSFAC; T04689; -
CC DR Genes; HGNC:7673; NCOR2.
CC MTM; 600848; -
CC DR InterPro; IPR001005; Myb DNA binding.
CC DR Pfam; PF00249; myb DNA-binding; 2.
CC DR SMART; SM00395; SANT; 2.
CC DR PROSITE; PS50090; MYB 3; 1.
CC KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
CC Coiled coil; Alternative splicing.
CC KM
CC FT DOMAIN 174 215
CC FT DOMAIN 254 312
CC FT DNA BIND 429 474
CC FT DNA BIND 613 657
CC FT DOMAIN 522 561
CC FT DOMAIN 778 820
CC FT DOMAIN 2139 2346
CC FT DOMAIN 2342 2346
CC FT DOMAIN 494 510
CC FT DOMAIN 682 685
CC FT DOMAIN 994 1002
CC FT DOMAIN 1384 1389
CC FT DOMAIN 1842 1846
CC FT DOMAIN 2479 2482
CC FT VARSPIC 1 1702
CC FT VARSPIC 2353 2398
CC FT CONFLICT 7 7
CC FT CONFLICT 295 295
CC FT CONFLICT 309 309
CC FT CONFLICT 352 352
CC MISSING (IN REF. 2).
CC MISSING (IN REF. 2).
CC MISSING (IN REF. 2).
CC MISSING (IN REF. 2).

FT CONFLICT 365 365 A -> P (IN REF. 2).
FT CONFLICT 612 613 SS -> EF (IN REF. 5).
FT CONFLICT 711 711 S -> T (IN REF. 2).
FT CONFLICT 724 740 MISSING (IN REF. 2).
FT CONFLICT 787 796 RRTSRAPIE -> PDIPAPTES (IN REF. 2).
FT CONFLICT 804 804 G -> L (IN REF. 2).
FT CONFLICT 814 814 S -> F (IN REF. 2).
FT CONFLICT 817 817 A -> S (IN REF. 2).
FT CONFLICT 889 889 G -> R (IN REF. 2).
FT CONFLICT 1023 1030 SRSPAPPA -> MEADPAHP (IN REF. 3).
FT CONFLICT 1034 1034 A -> AEKVPFPA (IN REF. 2).
FT CONFLICT 1894 1894 K -> T (IN REF. 4).
FT CONFLICT 1894 1894 P -> A (IN REF. 4).
FT CONFLICT 2494 2494
SQ SEQUENCE 2517 AA; 274031 MW; F5805C01761258C0 CRC64;

Query Match 16.4%; Score 72; DB 1; Length 2517;
Best local Similarity 35.8%; Pred. No. 77;
Matches 24; Conservative 5; Mismatches 28; Indels 10; Gaps 3;

Qy 25 PGAGPPRRK-----ESTPETE-EGAPTTSEKKRIRGKFPVYNLSEIQNVKSELK 78
Db 780 PGPTPRKTSRAPIEPTPASEATGAPT---PPAPSPSAPPVVKKEKEETAAAP 835
Qy 79 FVPKGEQ 85
Db 836 FVEEGEE 842

RESULT 12
TH1L METUA STANDARD; PRT; 381 AA.
ID TH1L METUA
AC 058341;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable thiamine biosynthesis protein th1l.
GN TH1L OR Mj0931.
OS Methanococcus jannaschii;
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96357999; PubMed=8688087;
RA Built C.U., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback I.R., Kelley J.M., Peterson J.D., Sadow F.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RT Science 273:1058-1073(1996).
RL
CC -1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE THIAZOLE MOIETY (BY
CC SIMILARITY).
CC -1- PATHWAY: Thiamine biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE TH1L FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67536; AAB98933.1; -
CC DR TIGR; Mj0931; -
CC DR InterPro; IPR004114; THUMP_dom.

DR InterPro: IPR003720; Thli.
 DR Pfam: PF02558; Thli; 1.
 DR Pfam: PF02926; THUMP; 1.
 DR TIGRPFAM: TIGR00342; Thli; 1.
 DR Thliane biosynthesis; Complete proteome.
 SQ SEQUENCE 381 AA; 43436 MW; 0A31F1069DA3357B CRC64;

Query Match 16.1%; Score 71; DB 1; Length 381;
 Best Local Similarity 28.4%; Pred. No. 12;
 Matches 23; Conservative 14; Mismatches 40; Indels 4; Gaps 3;

Qy 1 MSKOPISNVRAIQANINIPMGAPRRKSTPTEEGAPTSSEKK-PIPMKK 59
 Db 304 VASOTLNKLRISININPFI-LRPLGLDNDLVKIAK-ELGYEISTEKIKCYLPK 360
 Qy 60 PEPVNVLSIIONVKSLEKTV 80
 Db 361 HPKTIARPEEVKKIKKVKLV 381

RESULT 13
 GSRI_HUMAN STANDARD; PRT; 1509 AA.
 AC Q9NZM4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glioma tumor suppressor candidate gene 1 protein.
 GN GLTSCR1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=20175430; PubMed=10708517;
 RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thararajasingam U.,
 RA Porttler B.P., Ueki K., Billings S., Ramagwamy S., Mohrenweiser H.W.,
 RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
 RT "A transcrip map of the chromosome 19q-Arm glioma tumor suppressor
 RT region.";
 RL Genomics 64:44-50(2000).
 CC -1- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
 CC placenta, skeletal muscle, and pancreas, and at lower levels in
 CC lung, liver, and kidney.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF182077; AAF62874.1; -
 DR Genew; HGNC:4332; GLTSCR1.
 DR MIM; 605690; -
 FT DOMAIN 37 45 POLY-GLY.
 FT DOMAIN 884 889 POLY-PRO.
 FT DOMAIN 1214 1225 POLY-SER.
 FT DOMAIN 1282 1286 POLY-PRO.
 FT DOMAIN 1294 1304 POLY-PRO.
 SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;

Query Match 16.1%; Score 71; DB 1; Length 1509;
 Best Local Similarity 31.2%; Pred. No. 55;
 Matches 15; Conservative 9; Mismatches 20; Indels 4; Gaps 1;

Qy 18 IPMGAPRRKSTPTEEGAPTSSEKKPIPMKKPPGV 65
 Db 584 LPLGLDNDLVKIAK-ELGYEISTEKIKCYLPK 627

RESULT 14
 CES2_HUMAN STANDARD; PRT; 1484 AA.
 ID Q9BXF3; Q9C0C3; Q9EP58;
 AC Q9BXF3; Q9C0C3; Q9EP58;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cat eye syndrome critical region protein 2.
 GN CECR2 OR KIAA1740.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RX MEDLINE=21275466; PubMed=11381032;
 RA Foote T.K., Brinkman-Mills P., Banting G.S., Maier S.A., Riaz M.A.,
 RA Bridgland L.J., Hu S., Birren B., Minoshima S., Shimizu N., Pan H.,
 RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shaul S., Phan S., Yao Z.,
 RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,
 RA McDermid H.E.;
 RT "Analysis of the cat eye syndrome critical region in humans and the
 RT region of conserved syteny in mice: a search for candidate genes at
 RT or near the human chromosome 22 pericentromere.";
 RL Genome Res. 11:1053-1070(2001).
 RN [2]
 RP SEQUENCE OF 346-1484 FROM N.A. (ISOFORM A).
 RX TISSUE=Brain;
 RX MEDLINE=21082932; PubMed=11214970;
 RA Nagae T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 RN [3]
 RP SEQUENCE OF 168-1484 FROM N.A. (ISOFORM B), AND INTERACTION WITH
 RP LRPRC.
 RX TISSUE=Liver;
 RX MEDLINE=21686162; PubMed=11827465;
 RA Liu L., McKeahan W.L.;
 RT "Sequence analysis of LRPRC and its SEC1 domain interaction partners
 RT suggests roles in cytoskeletal organization, vesicular trafficking,
 RT nucleocytoplasmic shuttling, and chromosome activity.";
 RL Genomics 79:124-136(2002).
 CC -1- FUNCTION: May be involved through its interaction with LRPRC in
 CC the integration of cytoskeletal network with vesicular
 CC trafficking, nucleocytoplasmic shuttling, transcription, chromosome
 CC remodeling and cytokinesis.
 CC -1- SUBUNIT: Interacts with LRPRC.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here)
 CC and B/CECR2B; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in skeletal muscle, thymus,
 CC placenta and lung. Expressed at lower level in brain, heart,
 CC colon, spleen, kidney.
 CC -1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a
 CC developmental disorder associated with the duplication of a 2 Mb
 CC region of 22q11.2. Duplication usually takes in the form of a
 CC supernumerary bisected isodiscentric chromosome, resulting in
 CC four copies of the region (represents an inv dup(22)(q11)). CES is
 CC characterized clinically by the combination of coloboma of the
 CC iris and anal atresia with fistula, downslanting palpebral
 CC fissures, preauricular tags and/or pits, frequent occurrence of
 CC heart and renal malformations, and normal or near-normal mental
 CC development.
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

CC EMBL: AF316133; AAK15343.1; -
DR EMBL: AB051527; BAB21831.1; -
DR EMBL: AF411609; AAL07393.1; -
DR HSSP: Q92831; 1B91.
DR Genew; HGNC:1840; CECR2.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain.1.
DR PRINTS; PS00503; BROMODOMAIN.
DR SMART; SM00297; BROMO.1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
KW Alternative splicing.
FT DOMAIN 451 521 BROMODOMAIN.
FT DOMAIN 333 337 POLY-GLU.
FT DOMAIN 611 614 POLY-SER.
FT DOMAIN 1250 1253 POLY-PRO.
FT VASPLIC 291 318 MISSING (IN ISOFORM B).
FT VASPLIC 519 526 EYTKMSDN -> GKGRSLC (IN ISOFORM B).
FT VASPLIC 527 1484 MISSING (IN ISOFORM B).
FT CONFLICT 370 389 MISSING (IN REF. 2).
FT CONFLICT 1029 1029 C -> S (IN REF. 2).
FT CONFLICT 1045 1045 R -> W (IN REF. 2).
SQ SEQUENCE 1484 AA; 164214 MW; 049AA844E51AF63F CRC64;

```

```

Query Match 16.0%; Score 70.5; DB 1; Length 1484;
Best Local Similarity 29.6%; Pred. No. 60;
Matches 16; Conservative 13; Mismatches 20; Indels 5; Gaps 2;

```

```

Oy 16 INIPMGAFRPGAGQPPRRKSTETEGAPTSEKKPIPGMKPPGPVNLSE 69
Db 943 MSVTASAKPAPLGNPGR---APENSEAQEPENDQAPLPGLGEKP-PGVGRSE 991

```

RESULT 15

```

MACS CHICK STANDARD; PRT; 280 AA.
ID MACS_CHICK
AC P16527;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Myristoylated alanine-rich C-kinase substrate (MARCKS).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RA MEDLINE=90114197; PubMed=2608063;
RA Graff J.M., Stumpo D.J., Blackshear P.J.;
RT "Molecular cloning, sequence, and expression of a cDNA encoding the
RT chicken myristoylated alanine-rich C kinase substrate (MARCKS).";
RL Mol. Endocrinol. 3:1903-1906(1989).
CC -1- FUNCTION: MARCKS IS THE MOST PROMINENT CELLULAR SUBSTRATE FOR
CC PROTEIN KINASE C. THIS PROTEIN BINDS CALMODULIN, ACTIN, AND
CC SYNAPSIN. MARCKS IS A FILAMENTOUS (F) ACTIN CROSS-LINKING PROTEIN.
CC -1- PTM: PHOSPHORYLATION BY PKC DISPLACES MARCKS FROM THE MEMBRANE. IT
CC ALSO INHIBITS THE P-ACTIN CROSS-LINKING ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE MARCKS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31650; AAA48946.1; -
CC PIR: A41400;
CC InterPro: IPR002101; MARCKS.
CC Pfam; PF02063; MARCKS; 1.

```

```

DR PRINTS; PS00963; MARCKS.
DR PROSITE; PS00826; MARCKS_1; 1.
DR PROSITE; PS00827; MARCKS_2; 1.
KW Phosphorylation; Myristate; Calmodulin-binding; Actin-binding;
KW Membrane.
FT INIT MET 0 0
FT LIPID 1 1 MYRISTATE.
FT DOMAIN 116 140 CALMODULIN-BINDING (PSD).
FT MOD_RES 123 123 PHOSPHORYLATION (BY PKC).
FT MOD_RES 127 127 PHOSPHORYLATION (BY PKC).
FT MOD_RES 131 131 PHOSPHORYLATION (BY PKC).
FT MOD_RES 134 134 PHOSPHORYLATION (BY PKC).
SQ SEQUENCE 280 AA; 27597 MW; DFB4E9DC0B0839E CRC64;

```

```

Query Match 15.8%; Score 69.5; DB 1; Length 280;
Best Local Similarity 30.5%; Pred. No. 12;
Matches 18; Conservative 8; Mismatches 24; Indels 9; Gaps 1;

```

```

Oy 2 SKQPISTVRAIQANINIPMGAFRPGAGQPPRRKSTETEGAPTSEKKPIPGMKPF 60
Db 73 SSEPASEKKAEMAESTEPAS-----PAEGASPTTEGATPSSSSSTPKKKKRF 122

```

```

Search completed: April 9, 2003, 12:36:11
Job time : 14.6532 secs

```


GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:33:21 / Search time 21.5927 Seconds
(without alignments)
378,434 Million cell updates/sec

Title: US-09-647-019-2
Perfect score: 440
Sequence: 1 MSKQPSNVAIAQANINIPM.....NLSEIONVKSLEKFPKGEQ 85

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78.5	17.8	1240	1	VHWEV structural polypro
2	78.5	17.8	1242	2	S72350 structural polypro
3	78.5	17.8	1242	2	A56605 structural polypro
4	77	17.5	1870	2	S37671 MHC class III hist
5	77	17.5	1872	2	S36152 MHC class III hist
6	77	17.5	2142	2	B35098 MHC class II hist
7	76	17.3	226	2	A53273 structural polypro
8	75.5	17.2	1239	1	VHWEV structural polypro
9	74.5	16.9	994	2	S19595 chloride channel p
10	73	16.6	222	2	G86168 structural prote
11	73	16.6	1236	1	VHWEV structural polypro
12	72.5	16.5	445	2	B57742 cyclin Ib - maize
13	72.5	16.5	1241	2	S26373 genome polyprotein
14	72.5	16.5	1787	2	T20160 structural prote
15	72	16.4	376	2	S04497 surface antigen PA
16	72	16.4	428	2	T24769 hypothetical prote
17	72	16.4	1209	2	T00373 hypothetical prote
18	72	16.4	1385	2	T21706 energy transducer
19	71	16.1	239	2	H90847 hypothetical prote
20	71	16.1	239	2	G84705 conserved hypotet
21	71	16.1	381	1	C64416 l7123.15 protein - hu
22	70.5	16.0	2100	2	T38128 elastin
23	70	15.9	7962	2	I38346 myristylated alan
24	69.5	15.8	281	2	A41400 exocoxin 8 (import
25	69.5	15.8	356	2	A89807 Balbani ring 2.1
26	69	15.7	749	2	A45294 probable iron-sulp
27	69	15.7	880	2	B87222 membrane protein I
28	68.5	15.6	243	2	T45505 tonb protein - Ec
29	68	15.5	239	1	BVEC

30	68	15.5	253	2	JC2388 class II histocomp
31	68	15.5	440	2	T61183 transcripction fact
32	68	15.5	651	2	T40459 hypothetical lysin
33	67.5	15.3	568	2	G02753 testis specific ba
34	67.5	15.3	1113	2	T47381 hypothetical prote
35	67.5	15.3	1415	2	T21244 zyg-9 protein - Ca
36	67.5	15.2	485	2	T27147 hypothetical prote
37	67	15.2	598	2	A75531 hypothetical prote
38	67	15.2	1666	2	A48594 ekelemin - mouse
39	66.5	15.1	241	1	S64445 hypothetical prote
40	66.5	15.1	526	2	G69094 secretory protein
41	66.5	15.1	982	1	GNLJH2 pol polyprotein -
42	66.5	15.1	1152	2	A33183 microtubule-associ
43	66	15.0	240	2	S13257 tonb protein - Sal
44	66	15.0	252	2	A46505 SLA-DR4 (MHC Clas
45	66	15.0	401	2	D85090 hypothetical prote

ALIGNMENTS

RESULT 1

VHWEV structural polyprotein - eastern equine encephalomyelitis virus (strain VA3/78n Broeck N:Contains: 6K protein; coat protein C; membrane glycoprotein E1; membrane glycoprotein C;Species: eastern equine encephalomyelitis virus A:Note: host Equus caballus (domestic horse) C:Date: 30-Jun-1992 #sequence_reviation 30-Jun-1992 #text_change 16-Jul-1999 C:Accession: A39992 R:Weaver, S.C.; Scott, T.W.; Rico-Hesse, R. Virology 182, 774-784, 1991 A:Title: Molecular evolution of eastern equine encephalomyelitis virus in North America A:Reference number: A39992; MUID:9120727; PMID:2024496 A:Accession: A39992 A:Molecule type: genomic RNA A:Residues: 1-1240 <MEM> A:Cross-references: GB:M69094; NID:9323696; PIDN:AAA42980.1; PID:9323697 A:Note: the authors translated the codon AGC for residue 836 as Arg and GGU for residue C;Superfamily: togavirus structural polyprotein C:Keyword: coat protein; glycoprotein; polypeptide; transmembrane protein F:1-266/Product: coat protein C #status predicted <TM1> F:259-276/Domain: transmembrane #status predicted <TM1> F:261-323/Product: membrane glycoprotein E3 #status predicted <EG3> F:324-743/Product: membrane glycoprotein E2 #status predicted <EG2> F:695-712/Domain: transmembrane #status predicted <TM2> F:722-738/Domain: transmembrane #status predicted <TM3> F:744-799/Product: 6K protein #status predicted <KP6> F:781-799/Domain: transmembrane #status predicted <TM4> F:800-1240/Product: transmembrane glycoprotein E1 #status predicted <EG1> F:1212-1236/Domain: transmembrane #status predicted <TM5> (covalent) #status predicted F:49,271,625,638,834,933/Binding site: carbohydrate (Asn)

Query Match 17.8%; Score 78.5; DB 1; Length 1240; Best Local Similarity 33.3%; Pred. No. 1; Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVAIAQANINIPMGAFPGACQPPRRKSTPTEBGAFTTBKKKTPGMK 59
DB 40 IEDRRRIANILTKORAPNPAPGPAKRRKPAKPK---PAQAKKRRPPAPK 90

RESULT 2
S72350 structural polyprotein - eastern equine encephalomyelitis virus N:Contains: 6K protein; capsid protein; E1 protein; E2 protein; E3 protein C:Species: eastern equine encephalomyelitis virus C:Date: 04-May-1998 #sequence_reviation 15-May-1998 #text_change 26-Aug-1999 C:Accession: S72350 R:Weaver, S.C.; Hagenbaugh, A.; Bellw, L.A.; Netesov, S.V.; Volchkov, V.B.; Chang, G. Virology 197, 375-390, 1993 A:Title: A comparison of the nucleotide sequences of eastern and western equine encephalomyelitis virus A:Reference number: S72349; MUID:94025587; PMID:8105605 A:Accession: S72350

A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1242 <WEA>
 A:Cross-references: EMBL:U01034; NID:9393006; PID:AA53735.1; PID:9393008
 C:Superfamily: togavirus structural polypeptide

Query Match 17.8%; Score 78.5; DB 2; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 12;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNRAIQANINIMGAFRPGAGOPRRKSTPTEBGAFTTSEKKPIPGMKK 59
 Db 41 IEDLRRIANILTKQRAPNPAGPPAKKPKPKK--PAQTKKKRPPPAKK 91

RESULT 3
 A56605
 structural polypeptide - eastern equine encephalomyelitis virus (strain 4789)
 C:Species: eastern equine encephalomyelitis virus
 C>Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 26-Aug-1999
 C:Accession: A56605
 R:Weaver, S.C.; Hagenbaugh, A.; Bellow, L.A.; Calisher, C.H.
 Arch. Virol. 127, 305-314, 1992
 A:Title: Genetic characterization of an antigenic subtype of eastern equine encephalomyelitis virus
 A:Reference number: A56605; MUID:93090093; PMID:1280945
 A:Accession: A56605
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1242 <WEA>
 A:Cross-references: GB:120951, NID:9405814, PID:AA02897.1; PID:9305047
 A:Note: sequence inconsistent with nucleotide translation
 C:Superfamily: togavirus structural polypeptide

Query Match 17.8%; Score 78.5; DB 2; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 12;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNRAIQANINIMGAFRPGAGOPRRKSTPTEBGAFTTSEKKPIPGMKK 59
 Db 41 IEDLRRIANILTKQRAPNPAGPPAKKPKPKK--PAQTKKKRPPPAKK 91

RESULT 4
 S37671
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
 C:Species: Homo sapiens (man)
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
 C:Accession: S37671
 R:Bouguetel, L.
 submitted to the EMBL Data Library, August 1992
 A:Reference number: S37671
 A:Accession: S37671
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1870 <BOU>
 A:Cross-references: EMBL:Z15025; NID:929374; PID:929375
 C:Genetics:
 A:Map position: 6p21.3
 A:introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 65
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 17.5%; Score 77; DB 2; Length 1870;
 Best Local Similarity 30.4%; Pred. No. 26;
 Matches 24; Conservative 10; Mismatches 31; Indels 14; Gaps 3;

QY 3 KQPSNVRAIQANINIMGAFRPGAGOPRRKSTPTEBGAFTTSEKKPIPGMKK 55
 Db 886 KEETRAQLTGPEAGKRLP--ASRSAGAPPPRRSRRTETRWG--PRPGSSRRGIPPEEGAP 942
 QY 56 ----GMKKFPGPVNLSEI 70

Db 943 PRRAGPIKKPPPTKVEL 961

RESULT 5
 S36152
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - human
 C:Species: Homo sapiens (man)
 C>Date: 06-Jun-1995 #sequence_revision 17-Nov-1995 #text_change 15-Sep-2000
 C:Accession: S36152
 R:Lillis, F.J.M.; Bouguetel, L.; Prieur, S.; Caterina, D.; Primas, G.; Perrot, V.; Turk
 Nature Genet. 3, 137-145, 1993
 A:Title: Dense Alu clustering and a potential new member of the NF-kappaB family within
 A:Reference number: S36152; MUID:93272029; PMID:8499947
 A:Accession: S36152
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1872 <TRI>
 A:Cross-references: EMBL:Z15025
 A:Note: in the authors' translation residues 32-34 are shown after residue 4 and, conse
 C:Genetics:
 A:introns: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 6
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 17.5%; Score 77; DB 2; Length 1870;
 Best Local Similarity 30.4%; Pred. No. 26;
 Matches 24; Conservative 10; Mismatches 31; Indels 14; Gaps 3;

QY 3 KQPSNVRAIQANINIMGAFRPGAGOPRRKSTPTEBGAFTTSEKKPIPGMKK 55
 Db 887 KEETRAQLTGPEAGKRLP--ASRSAGAPPPRRSRRTETRWG--PRPGSSRRGIPPEEGAP 943

QY 56 ----GMKKFPGPVNLSEI 70
 Db 944 PRRAGPIKKPPPTKVEL 962

RESULT 6
 B35098
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [imported] - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Aug-1990 #sequence_revision 06-Nov-1992 #text_change 24-Aug-2001
 C:Accession: B35098
 R:Banerji, J.; Sande, J.; Strominger, J.L.; Spies, T.
 Proc. Natl. Acad. Sci. U.S.A. 87, 2374-2378, 1990
 A:Title: A gene pair from the human major histocompatibility complex encodes large prol
 A:Reference number: A35098; MUID:90192810; PMID:2156268
 A:Accession: B35098
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2142 <BAN>
 A:Cross-references: GB:M33509; NID:9179338; PID:AA35585.1; PID:9179339; GB:M31293
 C:Note: the authors translated the codon AGT for residue 97 as Gly
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 17.5%; Score 77; DB 2; Length 2142;
 Best Local Similarity 30.4%; Pred. No. 30;
 Matches 24; Conservative 10; Mismatches 31; Indels 14; Gaps 3;

QY 3 KQPSNVRAIQANINIMGAFRPGAGOPRRKSTPTEBGAFTTSEKKPIPGMKK 55
 Db 875 KEETRAQLTGPEAGKRLP--ASRSAGAPPPRRSRRTETRWG--PRPGSSRRGIPPEEGAP 931
 QY 56 ----GMKKFPGPVNLSEI 70
 Db 932 PRRAGPIKKPPPTKVEL 950

RESULT 7
 A53273
 MHC class II histocompatibility antigen DR alpha chain - horse (fragment)
 C:Species: Equus caballus (domestic horse)
 C>Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 21-Jan-2000


```
A:Reference number: A35587; MUID:86320369; PMID:3413072
A:Accession: A35587
A:Molecule type: genomic RNA
A:Residues: 1-1236 <HA>
A:Cross-references: GB:J03854; NID:g323728; PIDN:AAA2999.1; PTD:g323730
C:Superfamily: togavirus structural polyprotein
F:Keyword: coat protein; glycoprotein; polypeptide; transmembrane protein
F:1-259/Product: coat protein C #status predicted <CC>
F:260-319/Product: membrane glycoprotein E3 #status predicted <MG3>
F:320-742/Product: membrane glycoprotein E2 #status predicted <MG2>
F:684-702/Domin: transmembrane #status predicted <TN1>
F:719-733/Domin: transmembrane #status predicted <TN2>
F:743-797/Product: 6K protein #status predicted <KP>
F:777-793/Domin: transmembrane #status predicted <TN3>
F:798-1236/Product: membrane glycoprotein E1 #status predicted <MG1>
F:1206-1227/Domin: transmembrane #status predicted <TN5>
F:50_270_515_637_724_936_1042/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

Query Match
Best Local Similarity 16.6%; Score 73; DB 1; Length 1236;
Matches 16; Conservative 14; Mismatches 24; Indels 6; Gaps 1;

Cyclin Db - maize cyclin B57742
C:Species: Zea mays (maize)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 16-Jul-1999
C:Accession: B57742
R:Reaudin, J.P.; Colasanti, J.; Rime, H.; Yuan, Z.; Sundaresan, V.
Proc. Natl. Acad. Sci. U.S.A. 91, 7375-7379, 1994
A>Title: Cloning of four cyclins from maize indicates that higher plants have three str
A:Reference number: A57742; MUID:94316698; PMID:8041798
A:Accession: B57742
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-445 <REN>
A:Cross-references: GB:U10076; GB:U10077; GB:U10078; NID:g516551; PIDN:AAA20238.1; PTD:

Query Match
Best Local Similarity 16.5%; Score 72.5; DB 2; Length 445;
Matches 26; Conservative 9; Mismatches 17; Indels 45; Gaps 5;

Oy 4 OPISVRAIQ-----ANINPMGAFRRACOPPPRKSTPTTEGA-----44
 |||:|||:
Db 88 QPVA-AEAVTKARKYTTKNVP---RPAGQAP--KENKKPSASGGAAASGRSVGNRR 140
 |||||
Oy 45 -----PTTSEKKPIPGMKKPFPG 62
 |::|||:|:
Db 141 KKPACTLTSTVLARSKAACPTEKEPEIEDIKFDG 177
 |::|||:|:

RESULT 13
S26373 genome polyprotein - eastern equine encephalomyelitis virus
N:Contains: 6K protein; capsid protein C; envelope protein E1; envelope protein E2; env
C:Species: eastern equine encephalomyelitis virus
C>Date: 06-Jan-1994 #sequence_revision 01-Nov-1996 #text_change 26-Aug-1999
C:Accession: S26373
R:Volchikov, V.E.; Volchkova, V.A.; Netesov, S.V.
Mol. Gen. Microbiol. Virol. 5, 8-15, 1991
A>Title: Complete nucleotide sequence of the eastern equine encephalomyelitis virus genc
A:Reference number: S26369; MUID:91375524; PMID:1896061
A:Accession: S26373
A:Molecule type: mRNA
A:Residues: 1-1241 <VOL>
A:Cross-references: EMBL:X63135; NID:g59185; PIDN:CAA44845.1; PTD:g59186

```

A>Note: sequence could not be checked because of bad print in paper
C:Superfamily: togavirus structural polypeptide
C:Keywords: capsid protein; envelope protein; glycoprotein; polypeptide
F:1-260/Product: capsid protein C #status predicted <CAP>
F:161-323/Product: envelope protein E3 #status predicted <EP3>
F:324-743/Product: envelope protein E2 #status predicted <EP2>
F:744-800/Product: 6K protein #status predicted <6KP>
F:801-1241/Product: envelope protein E1 #status predicted <EP1>

Query Match
Best Local Similarity 16.5%; Score 72.5; DB 2; Length 1241;
Matches 17; Conservative 12; Mismatches 25; Indels 7; Gaps 1;

QY 6 ISNVRAIQANINIPMGAFPRGAGPPRRKESRP-----ETEEGAPITSEKKPIPGMK 58
DB 41 IEDLRSLNLTLLKQRPNPAGPPAKRRKPPAPSLRRKKRPPPPAKKQRRKPPGR 100
QY 59 K 59
DB 101 Q 101

RESULT 14
T20160
Hypothetical protein T14G8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
R:McMurray, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19231
A:Accession: T20160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <MTL>
A:Cross-references: EMBL:Z67881; PIDN:CAA91798.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone C52G5
R:Matthews, P.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19955
A:Accession: T24924
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <MT2>
A:Cross-references: EMBL:Z67884; PIDN:CAA91810.1; GSPDB:GN00028; CESP:T14G8.1
C:Genetics:
A:Gene: CESP:T14G8.1
A:Map position: X
A:Introns: 112/3; 453/3; 597/3; 815/2; 1258/3; 1682/2; 1709/3; 1764/1

Query Match
Best Local Similarity 16.5%; Score 72.5; DB 2; Length 1787;
Matches 20; Conservative 3; Mismatches 11; Indels 15; Gaps 2;

QY 35 ESTPETER-GAPITSEKKPIPGMK-----FPGPVNIS 68
DB 36 EEEVETESQGVPTTSEKKPPPKKKGGKSSKKKNNCDYDPYKST 84

RESULT 15
S04497
surface antigen PAg - Streptococcus sobrinus (fragment)
N:Alternate names: surface antigen Spa
C:Species: Streptococcus sobrinus
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 07-Feb-1997
C:Accession: S04497
R:Takahashi, I.; Okahashi, N.; Saeakawa, C.; Yoshikawa, M.; Hamada, S.; Koga, T.
FEBS Lett. 249, 383-388, 1989
A:Title: Homology between surface protein antigen genes of Streptococcus sobrinus and S.
A:Reference number: S04497; MUID:89250018; PMID:2661267
A:Accession: S04497
A:Status: not compared with conceptual translation

```

A: Molecule type: DNA
A: Residues: 1-376 <TAK>
C: Genetices:
A: Gene: pag
C: Superfamily: surface antigen gpap
C: keywords: surface antigen

Query Match 16.4%; Score 72; DB 2; Length 376;
Best Local Similarity 26.9%; Pred. No. 14;
Matches 21; Conservative 9; Mismatches 30; Indels 18; Gaps 4;

QY 2 SKQPIINV--RAIQANINIFMGAFRPGAGOPPRKKESTPETEGAPTSEKKPI----- 54
DB 264 SKTPDQNIIPDKVPVPTVEVEK-BLEPGTSEVNEKEPTP--POSTPDQEEPTKPVPEPSYQ 320
QY 55 -----PGMKKPPGPV 64
DB 321 SLPTPPVAPTYEKVGPV 338

Search completed: April 9, 2003, 12:39:36
Job time : 25.5927 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 9, 2003, 12:31:56 ; Search time 43.5282 Seconds
(without alignments)
402.360 Million cell updates/sec

Title: US-09-647-019-2

Perfect score: 440
Sequence: 1 MSKQPSNVRATQANINIM.....NLSEIQVKSLEKFPVKGRO 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	440	100.0	85	11 Q9ER98	Q9ER98 mus musculus
2	410	93.2	85	11 Q925F0	Q925F0 ratu
3	326.5	74.2	91	13 Q90258	Q90258 xenopus lae
4	81.5	18.5	976	6 Q9MZT1	Q9MZT1 canis fami
5	79.5	18.1	1242	12 Q9PZX7	Q9PZX7 eastern equ
6	78.5	17.8	1242	12 Q9PZX6	Q9PZX6 eastern equ
7	78.5	17.8	1242	12 Q9PZX5	Q9PZX5 eastern equ
8	78.5	17.8	1242	12 Q9PZX4	Q9PZX4 eastern equ
9	78.5	17.8	1242	12 Q9PZX3	Q9PZX3 eastern equ
10	78.5	17.8	1242	12 Q9PZX2	Q9PZX2 eastern equ
11	78.5	17.8	1242	12 Q9PZX1	Q9PZX1 eastern equ
12	78.5	17.8	1242	12 Q9PZX0	Q9PZX0 eastern equ
13	78.5	17.8	1242	12 Q9PZX9	Q9PZX9 eastern equ
14	78.5	17.8	1242	12 Q9PZX8	Q9PZX8 eastern equ
15	78.5	17.8	1242	12 Q9PZX7	Q9PZX7 eastern equ
16	78.5	17.8	1242	12 Q9PZX6	Q9PZX6 eastern equ

17	78.5	17.8	1242	12 Q98794	Q98794 eastern equ
18	78.5	17.8	1242	12 Q98795	Q98795 eastern equ
19	78.5	17.8	1242	12 Q98796	Q98796 eastern equ
20	78.5	17.8	1242	12 Q98797	Q98797 eastern equ
21	78.5	17.8	1242	12 Q98798	Q98798 eastern equ
22	78.5	17.8	1242	12 Q98799	Q98799 eastern equ
23	78.5	17.8	1242	12 Q98359	Q98359 eastern equ
24	77.5	17.6	5085	11 Q9JX56	Q9JX56 ratu
25	76.5	17.4	182	2 Q9ADV0	Q9ADV0 ehrlchia c
26	76	17.3	226	2 Q90489	Q90489 equus caball
27	76	17.3	1404	4 Q9BX49	Q9BX49 homo sapien
28	76	17.3	1404	4 Q92954	Q92954 homo sapien
29	75.5	17.2	682	5 Q9U010	Q9U010 drosophila
30	75.5	17.2	745	5 Q61458	Q61458 drosophila
31	75.5	17.2	771	5 Q76259	Q76259 drosophila
32	75.5	17.2	771	5 Q62609	Q62609 drosophila
33	74	16.8	218	7 Q19433	Q19433 felis silve
34	74	16.8	254	7 Q19432	Q19432 felis silve
35	74	16.8	254	7 Q19434	Q19434 felis silve
36	74	16.8	411	10 Q9LFM5	Q9LFM5 arabidopsis
37	74	16.8	1242	12 Q9PZX8	Q9PZX8 eastern equ
38	74	16.8	1242	12 Q9PZX7	Q9PZX7 eastern equ
39	73.5	16.7	1242	12 Q98678	Q98678 eastern equ
40	73.5	16.7	1242	12 Q9PZX6	Q9PZX6 eastern equ
41	73	16.6	222	10 Q9ZMA4	Q9ZMA4 arabidopsis
42	73	16.6	297	12 Q8V719	Q8V719 tt virus. o
43	73	16.6	437	2 Q9P1V8	Q9P1V8 streptomyce
44	72.5	16.5	266	6 Q9SKV6	Q9SKV6 bos taurus
45	72.5	16.5	445	10 Q41733	Q41733 zea mays (m

ALIGNMENTS

RESULT 1
ID Q9ER98 PRELIMINARY; PRT; 85 AA.

AC Q9ER98;
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 19, Last annotation update)
DE Stretch responsive muscle (X-chromosome) (SMX protein)
DE (Muscle-specific protein CSL)
GN SMX OR SRMX OR CSL
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SKELTAL MUSCLE;
RX MEDLINE=2125047; PubMed=1101441;
RA Kemp T.J., Sadusky T.J., Simon M., Brown R., Eastwood R.,
RA Sasso D.A., Coulton G.R.;
RT "Identification of a Novel Stretch-Responsive Skeletal Muscle Gene
RT (smx).";
RE Genomics 72:260-271 (2001).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065879; PubMed=10598820;
RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.;
RT "Identification, mapping, and genomic structure of a novel X-
RT chromosomal human gene (SMX) encoding a small muscular protein.";
RL Hum. Genet. 105:506-512 (1999).
RL [3]
RP SEQUENCE FROM N.A.
RA Patzak D.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=2125706; PubMed=11381084;
RA Palmer S., Groves N., Schindeler A., Yeoh T., Biben C., Wang C.-C.,

RA Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,
 RA Mohun T., Harvey R.P.;
 RT "The Small Muscle-specific Protein Csl Modifies Cell Shape and
 RT Promotes Myocyte Fusion in an Insulin-like Growth Factor 1-dependent
 RT Manner.";
 RL J. Cell Biol. 153:985-998(2001).
 DR EMBL; AJ245772; CAC08493.1; -;
 DR EMBL; AF364070; AAK50398.1; -;
 DR EMBL; AY026524; AAK07682.1; -;
 DR MGI; 1913356; SMPX.
 SQ SEQUENCE 85 AA; 9253 MW; 43863840A65DA6BC CRC64;

Query Match 100.0%; Score 440; DB 11; Length 85;
 Best Local Similarity 100.0%; Pred. No. 1.2e-40;
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 60
 DB 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 60

QY 61 PGPVNLSEIQNVKSELKFPKGEQ 85
 DB 61 PGPVNLSEIQNVKSELKFPKGEQ 85

RESULT 2

Q925FO PRELIMINARY; PRT; 85 AA.

ID Q925FO
 AC Q925FO
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE SMPX protein.
 GN SMPX.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20065879; PubMed=10598820;
 RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.;
 RT "Identification, mapping, and genomic structure of a novel X-
 RT chromosome human gene (SMPX) encoding a small muscular protein.";
 RL Hum. Genet. 105:506-512(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Patzak D.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF364071; AAK50399.1; -;
 SQ SEQUENCE 85 AA; 9120 MW; 52654F8C790C932C CRC64;

Query Match 93.2%; Score 410; DB 11; Length 85;
 Best Local Similarity 91.8%; Pred. No. 2.2e-37;
 Matches 78; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 60
 DB 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 60

QY 61 PGPVNLSEIQNVKSELKFPKGEQ 85
 DB 61 PGPVNLSEIQNVKSELKFPKGEQ 85

RESULT 3

Q90258 PRELIMINARY; PRT; 91 AA.

ID Q90258
 AC Q90258
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Chisel

CN Csl.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21275706; PubMed=11381084;
 RA Palmer S., Groves N., Schindeler A., Yeoh T., Bibben C., Wang C.-C.,
 RA Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,
 RA Mohun T., Harvey R.P.;
 RT "The Small Muscle-specific Protein Csl Modifies Cell Shape and
 RT Promotes Myocyte Fusion in an Insulin-like Growth Factor 1-dependent
 RT Manner.";
 RL J. Cell Biol. 153:985-998(2001).
 DR EMBL; AF343894; AAK71068.1; -;
 SQ SEQUENCE 91 AA; 10006 MW; BD2BA90B82B3846C CRC64;

Query Match 74.2%; Score 326.5; DB 13; Length 91;
 Best Local Similarity 72.2%; Pred. No. 3e-28;
 Matches 65; Conservative 6; Mismatches 14; Indels 5; Gaps 1;

QY 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 55
 DB 1 MSKOPISNVRAIOANINIPMGAFRPGAGOPPRKSTPTEEGAPTSEKKPIPMKKF 60

QY 56 GKKKPGPVNLSEIQNVKSELKFPKGEQ 85
 DB 61 GAVKLPGAPNLSEIQNVKSELKFPKGEQ 90

RESULT 4

Q9MZT1 PRELIMINARY; PRT; 976 AA.

ID Q9MZT1
 AC Q9MZT1
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Skeletal muscle chloride channel ClC-1.
 GN CLCN1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-BREED MINIATURE SCHNAUZER;
 RX MEDLINE=99379598; PubMed=10452529;
 RA Rhodes T.H., Vite C.H., Giger U., Patterson D.F., Pahlke C.,
 RA George A.L., Jr.;
 RT "A missense mutation in canine ClC-1 causes recessive myotonia
 RT congenita in the dog.";
 RL FEBS Lett. 456:54-58(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-BREED MINIATURE SCHNAUZER;
 RX ROGERS C.S., George A.L., Jr., Rhodes T.H.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF162445; AAF82606.1; -;
 DR InterPro; IPR006044; CBS domain.
 DR InterPro; IPR001807; Cl-channel_volt.
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00571; CBS; 2.
 DR Pfam; PF00654; voltage_CIC; 1.
 DR PRINTS; PR00762; CLCHANNEL.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
 FT VARIANT 268 T->M.
 SQ SEQUENCE 976 AA; 108053 MW; 8F85593E7C858F6E CRC64;

Query Match 18.5%; Score 81.5; DB 6; Length 976;
 Best Local Similarity 34.7%; Pred. No. 2.3;
 Matches 26; Conservative 7; Mismatches 27; Indels 15; Gaps 4;


```

QY 19 PMGAFR-----PGAGPPRRKSTPTEBGAFTTSEKKIPGMKKRPVNLSEI 70
DB 878 PLASFSTSTTRKNGGPPPTTAWSLPEOGTAPASPEPPAPSPS---PAPL--LSEA 931
QY 71 -QNVKSELKFPVCKE 84
DB 932 PAYEGELELELGE 946

RESULT 5
Q9PZX7 PRELIMINARY; PRT; 1242 AA.
AC Q9PZX7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
OC Viruses; serina positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA38-MASS;
RA Brault A.C., Powers A.M., Kang W., Teoh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159550; AF04792.1; -
DR HSSP; P03315; IYCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01569; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR KMW
SQ SEQUENCE 1242 AA; 137667 MW; DF23D3631A6CE51A CRC64;

Query Match 18.1%; Score 79.5; DB 12; Length 1242;
Best Local Similarity 33.3%; Pred. No. 5;
Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVRALQANINIPMGAFRPGAGPPRRKSTPTEBGAFTTSEKKIPGMKK 59
DB 41 IEDLRISANLTLKORANPNPAGPPAKKKAPKPK--PAQAKKKRPPPAKK 91

RESULT 6
Q9PZX6 PRELIMINARY; PRT; 1242 AA.
AC Q9PZX6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
OC Viruses; serina positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LA50-ARTH67;
RA Brault A.C., Powers A.M., Kang W., Teoh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America."

```

```

RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159551; AF04793.1; -
DR HSSP; P03315; IYCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01569; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR KMW
SQ SEQUENCE 1242 AA; 137537 MW; 5C6D7A2060F20324 CRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
Best Local Similarity 33.3%; Pred. No. 6.5;
Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVRALQANINIPMGAFRPGAGPPRRKSTPTEBGAFTTSEKKIPGMKK 59
DB 41 IEDLRISANLTLKORANPNPAGPPAKKKAPKPK--PAQAKKKRPPPAKK 91

RESULT 7
Q9PZX5 PRELIMINARY; PRT; 1242 AA.
AC Q9PZX5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
OC Viruses; serina positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS83-4789;
RA Brault A.C., Powers A.M., Kang W., Teoh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159552; AF04794.1; -
DR HSSP; P03315; IYCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01569; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR KMW
SQ SEQUENCE 1242 AA; 137523 MW; 5C8B3BA44805B799 CRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
Best Local Similarity 33.3%; Pred. No. 6.5;
Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVRALQANINIPMGAFRPGAGPPRRKSTPTEBGAFTTSEKKIPGMKK 59
DB 41 IEDLRISANLTLKORANPNPAGPPAKKKAPKPK--PAQAKKKRPPPAKK 91

RESULT 8
Q9PZX4 PRELIMINARY; PRT; 1242 AA.
AC Q9PZX4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
OC Viruses; serina positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OC NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LA50-ARTH67;
RA Brault A.C., Powers A.M., Kang W., Teoh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America."

```

DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Structural polypeptide.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphaviruses;
 NCBI_TaxID=11021;
 RN
 RP
 RC STRAIN=TX91-V1-7164;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF159553; AAF04795.1; -
 DR HSSP; P03315; 1VCP.
 DR MEROPS; S03.001; -
 DR InterPro; IPR001836; Alpha core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR00936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha core; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR Pfam; PF00943; Alpha_E2_glycop; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Polypeptide.
 SQ SEQUENCE 1242 AA; 137594 MW; 27632P943DBCRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIQANINIPMGAFRPGAGOPPRKKESTPETEGAPTSEKKPIPGMKK 59
 ID IEDLRISINLTIKORAPNPAGPPAKKKKPAKPK---PAQAKKKRPPPAKK 91

RESULT 9
 Q9PEX3 PRELIMINARY; PRT; 1242 AA.
 AC Q9PEX3;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Structural polypeptide.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphaviruses;
 NCBI_TaxID=11021;
 RN
 RP
 RC STRAIN=FL93-939;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF159554; AAF04796.1; -
 DR HSSP; P03315; 1VCP.
 DR InterPro; IPR001836; Alpha core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR00936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha core; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR Pfam; PF00943; Alpha_E2_glycop; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Polypeptide.

KW Polypeptide.
 SQ SEQUENCE 1242 AA; 137613 MW; 6F474E82A91FP4CD CRC64;
 Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIQANINIPMGAFRPGAGOPPRKKESTPETEGAPTSEKKPIPGMKK 59
 ID IEDLRISINLTIKORAPNPAGPPAKKKKPAKPK---PAQTKKKRPPPAKK 91

RESULT 10
 Q9PEX2 PRELIMINARY; PRT; 1242 AA.
 AC Q9PEX2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 17, Last annotation update)
 DE Structural polypeptide.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphaviruses;
 NCBI_TaxID=11021;
 RN
 RP
 RC STRAIN=TX95-FV5-2547;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF159555; AAF04797.1; -
 DR HSSP; P03315; 1VCP.
 DR MEROPS; S03.001; -
 DR InterPro; IPR001836; Alpha core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR00936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha core; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR Pfam; PF00943; Alpha_E2_glycop; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Polypeptide.
 SQ SEQUENCE 1242 AA; 137604 MW; 23BB89B5786F3B0E CRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIQANINIPMGAFRPGAGOPPRKKESTPETEGAPTSEKKPIPGMKK 59
 ID IEDLRISINLTIKORAPNPAGPPAKKKKPAKPK---PAQAKKKRPPPAKK 91

RESULT 11
 Q9PEX1 PRELIMINARY; PRT; 1242 AA.
 AC Q9PEX1;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Structural polypeptide.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphaviruses;
 NCBI_TaxID=11021;
 RN
 RP
 RC STRAIN=FL96-14834;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF159556; AAF04798.1; -
 DR HSSP; P03315; 1VCP.
 DR InterPro; IPR001836; Alpha core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR00936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha core; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR Pfam; PF00943; Alpha_E2_glycop; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 DR Polypeptide.

RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RT Viruses from North, Central and South America."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF159556; AAF04798.1; -.
 DR HSSP: P03315; 1VCP.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00944; Alpha_core; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KM Polyprotein.
 SQ SEQUENCE 1242 AA; 137637 MW; 4235EFD1FD501F7 CRC64;
 Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIOANINIPMGAFRPGAGOPRRKKESTPETEGAPTSEKKKPIPGMK 59
 DB 41 IEDLRISIANLTIKORANPNPAGPPAKKKKPAKPK---PAQAKKKRPPPAKK 91
 RESULT 12
 Q9P2X0 PRELIMINARY; PRT; 1242 AA.
 AC Q9P2X0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Structural polyprotein.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
 OS virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GAG1-FOREE;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RT Viruses from North, Central and South America."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF159557; AAF04799.1; -.
 DR HSSP: P03315; 1VCP.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00944; Alpha_core; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KM Polyprotein.
 SQ SEQUENCE 1242 AA; 137569 MW; BF2D08B2EBF5468 CRC64;
 Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIOANINIPMGAFRPGAGOPRRKKESTPETEGAPTSEKKKPIPGMK 59
 DB 41 IEDLRISIANLTIKORANPNPAGPPAKKKKPAKPK---PAQAKKKRPPPAKK 91

RESULT 13
 Q9P2W9 PRELIMINARY; PRT; 1242 AA.
 AC Q9P2W9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Structural polyprotein.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
 OS virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MK97-1076;
 RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine Encephalitis
 RT Viruses from North, Central and South America."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF159558; AAF04800.1; -.
 DR HSSP: P03315; 1VCP.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00944; Alpha_core; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KM Polyprotein.
 SQ SEQUENCE 1242 AA; 137640 MW; F728C704E178E99B CRC64;
 Query Match 17.8%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 33.3%; Pred. No. 6.5;
 Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;
 QY 6 ISNVAIOANINIPMGAFRPGAGOPRRKKESTPETEGAPTSEKKKPIPGMK 59
 DB 41 IEDLRISIANLTIKORANPNPAGPPAKKKKPAKPK---PAQAKKKRPPPAKK 91
 RESULT 14
 Q88790 PRELIMINARY; PRT; 1242 AA.
 AC Q88790;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Structural polyprotein.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
 OS virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 NC NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NORTH AMERICAN ANTIGENIC VARIETY;
 RX MEDLINE=94025587; Pubmed=8105605;
 RA Weaver S.C., Hagenbaugh A., Bellew L., Netesov S.V., Volchokov V.E.,
 RA Chang G.J., Clarke D.K., Goussier L., Scott T.W., Trent D.W.,
 RA Holland J.J.;
 RT "A comparison of the nucleotide sequences of eastern and western
 RT equine encephalomyelitis viruses with those of other alphaviruses and
 RT related RNA viruses."
 RL Virology 197:375-390(1993).
 DR EMBL: U01034; AAC53735.1; -.
 DR HSSP: P03315; 1VCP.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR002548; Alpha_E1_glycop.

DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polypeptide.
FT CHAIN 1 261 CAPSID.
FT CHAIN 262 324 E3.
FT CHAIN 325 744 E2.
FT CHAIN 745 801 6K.
FT CHAIN 802 1242 E1.
SQ SEQUENCE 1242 AA; 137627 MW; BCALB498B4ABD27 CRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
Best Local Similarity 33.3%; Pred. No. 6.5;
Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVRAIQANINIMGAFRPGAGPPRRKESTPETEGAPTTSEKKRIPGMKK 59
DB 41 IEDLRRIANLTLKQAPNPPAGPPAKKKKAPKPK--PAQAKKKRPPPAKK 91

RESULT 15

088792 PRELIMINARY; PRT; 1242 AA.
AC 088792;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DECUR;
RX MEDLINE=94076404; PubMed=8254725;
RA Weaver S.C., Hagenbaugh A., Bellew L.A., Gousset L., Mallampalli V.,
Holland J.J., Scott T.W.;
RT "Evolution of alphaviruses in the eastern equine encephalomyelitis
RT complex.";
RL J. Virol. 68:158-169(1994).
DR EMBL; U01552; AAC53755.1; -.
DR HSSP; P03315; 1VCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polypeptide.
FT CHAIN 1 261 CAPSID.
FT CHAIN 262 321 E3.
FT CHAIN 322 744 E2.
FT CHAIN 745 772 6K.
FT CHAIN 773 1242 E1.
SQ SEQUENCE 1242 AA; 137523 MW; 08B5FF55D940DE CRC64;

Query Match 17.8%; Score 78.5; DB 12; Length 1242;
Best Local Similarity 33.3%; Pred. No. 6.5;
Matches 18; Conservative 12; Mismatches 21; Indels 3; Gaps 1;

QY 6 ISNVRAIQANINIMGAFRPGAGPPRRKESTPETEGAPTTSEKKRIPGMKK 59
DB 41 IEDLRRIANLTLKQAPNPPAGPPAKKKKAPKPK--PAQAKKKRPPPAKK 91

DB 41 IEDLRRIANLTLKQAPNPPAGPPAKKKKAPKPK--PAQAKKKRPPPAKK 91

Search completed: April 9, 2003, 12:38:25
Job time : 46.5282 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:38:46 ; Search time 18.8508 Seconds
(without alignments)
275.668 Million cell updates/sec

Title: US-09-647-019-2

Perfect score: 440
Sequence: 1 MSKQPSNVRAIQANINPM.....NLSEIQVKSLEKVPKGRQ 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodaca/2/pubpaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodaca/2/pubpaa/PCR_NEW_PUB pep:*
- 3: /cgn2_6/ptodaca/2/pubpaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodaca/2/pubpaa/US07_PUBCOMB pep:*
- 5: /cgn2_6/ptodaca/2/pubpaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodaca/2/pubpaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodaca/2/pubpaa/US07_PUBCOMB pep:*
- 8: /cgn2_6/ptodaca/2/pubpaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodaca/2/pubpaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodaca/2/pubpaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodaca/2/pubpaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodaca/2/pubpaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodaca/2/pubpaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodaca/2/pubpaa/US60_PUBCOMB pep:*

Pred: No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	17.3	941	12 US-10-124-557-14	Sequence 14, Appl
2	76	17.3	1022	12 US-10-124-557-84	Sequence 84, Appl
3	76	17.3	1038	12 US-10-124-557-74	Sequence 74, Appl
4	76	17.3	1049	12 US-10-124-557-58	Sequence 58, Appl
5	76	17.3	1140	12 US-10-124-557-104	Sequence 104, Appl
6	76	17.3	1270	12 US-10-124-557-44	Sequence 44, Appl
7	76	17.3	1311	12 US-10-124-557-42	Sequence 42, Appl
8	76	17.3	1313	12 US-10-124-557-142	Sequence 142, Appl
9	76	17.3	1314	12 US-10-124-557-50	Sequence 50, Appl
10	76	17.3	1320	12 US-10-124-557-60	Sequence 60, Appl
11	76	17.3	1354	12 US-10-124-557-48	Sequence 48, Appl
12	76	17.3	1361	12 US-10-124-557-40	Sequence 40, Appl
13	76	17.3	1361	12 US-10-124-557-52	Sequence 52, Appl
14	76	17.3	1404	12 US-10-124-557-62	Sequence 62, Appl
15	76	17.3	1404	12 US-10-124-557-62	Sequence 62, Appl
16	70.5	16.0	363	9 US-08-781-986A-5237	Sequence 5237, App
17	70.5	16.0	603	9 US-09-764-868-705	Sequence 705, App
18	66.5	15.1	441	12 US-10-081-281-115	Sequence 115, App

20	66.5	15.1	523	10 US-09-799-777-76	Sequence 76, Appl
21	65.5	14.9	196	10 US-09-858-664A-16	Sequence 16, Appl
22	65.5	14.9	616	9 US-10-243-735-4	Sequence 4, Appl
23	65.5	14.9	774	9 US-10-162-706-5	Sequence 5, Appl
24	64.5	14.7	603	10 US-09-906-779-4	Sequence 4, Appl
25	64.5	14.7	667	10 US-09-896-852-55	Sequence 55, Appl
26	64.5	14.7	1274	9 US-10-020-215-2	Sequence 2, Appl
27	64.5	14.7	1478	10 US-09-801-368-52	Sequence 52, Appl
28	64	14.5	212	10 US-09-733-507-12	Sequence 12, Appl
29	64	14.5	3014	10 US-09-737-149-2	Sequence 2, Appl
30	63.5	14.4	247	9 US-09-981-353-95	Sequence 95, Appl
31	63.5	14.4	280	10 US-09-925-300-1411	Sequence 1411, App
32	63	14.3	275	10 US-09-925-300-1491	Sequence 1491, App
33	63	14.3	501	9 US-09-738-626-6084	Sequence 6084, App
34	63	14.3	771	9 US-09-982-107-8	Sequence 8, Appl
35	63	14.3	892	10 US-09-205-658-42	Sequence 42, Appl
36	63	14.3	892	10 US-09-844-353A-42	Sequence 42, Appl
37	63	14.3	2507	9 US-09-819-104A-2	Sequence 2, Appl
38	62.5	14.2	196	12 US-10-081-281-117	Sequence 117, App
39	62.5	14.2	234	12 US-10-081-281-119	Sequence 119, App
40	62.5	14.2	330	10 US-09-815-837-94	Sequence 94, Appl
41	62.5	14.2	547	10 US-09-815-837-90	Sequence 90, Appl
42	62.5	14.2	551	10 US-09-815-837-93	Sequence 93, Appl
43	62.5	14.2	553	10 US-09-815-837-50	Sequence 50, Appl
44	62.5	14.2	555	10 US-09-815-837-71	Sequence 71, Appl
45	62.5	14.2	559	10 US-09-815-837-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1
US-10-124-557-14
Sequence 14, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Geeher, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESSES:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Query Match 17.3%; Score 76; DB 12; Length 941;
Best Local Similarity 42.2%; Pred. No. 9.7;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPTTEGAPTSEE-----KKPIPKMKPPGP 63
DB 257 PKPPTTPKKEPAPTTKEPAPTTKEPAPTAAPKKAPPTPKKEPAP 301

RESULT 2

US-10-124-557-84
Sequence 84, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gesner, Rodney M.
Hewick, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 84:
US-10-124-557-84

Query Match 17.3%; Score 76; DB 12; Length 1022;
Best Local Similarity 42.2%; Pred. No. 11;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPTTEGAPTSEE-----KKPIPKMKPPGP 63
DB 338 PKPPTTPKKEPAPTTKEPAPTTKEPAPTAAPKKAPPTPKKEPAP 382

RESULT 3

US-10-124-557-74
Sequence 74, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:

APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Gesner, Rodney M.
Hewick, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-10-124-557-74

Query Match 17.3%; Score 76; DB 12; Length 1038;
Best Local Similarity 42.2%; Pred. No. 11;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPTTEGAPTSEE-----KKPIPKMKPPGP 63
DB 322 PKPPTTPKKEPAPTTKEPAPTTKEPAPTAAPKKAPPTPKKEPAP 366

RESULT 4

US-10-124-557-58

Sequence 58, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Geesner, Thomas G.
Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58
Query Match 17.3%; Score 76; DB 12; Length 1049;
Best Local Similarity 42.2%; Pred. No. 11;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
QY 25 PGAGOPRRKKESTPTEGAPTTSE-----KKPIGKMKKFPGP 63
DB 365 PKKPTPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 409
RESULT 5
US-10-124-557-104
Sequence 104, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Geesner, Thomas G.
Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104
Query Match 17.3%; Score 76; DB 12; Length 1140;
Best Local Similarity 42.2%; Pred. No. 12;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
QY 25 PGAGOPRRKKESTPTEGAPTTSE-----KKPIGKMKKFPGP 63
DB 456 PKKPTPTPKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPAPTTKKPA 500
RESULT 6
US-10-124-557-44
Sequence 44, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacob, Kenneth
Geesner, Thomas G.
Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 104:
US-10-124-557-104

LENGTH: 1313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142

Query Match 17.3% Score 76; DB 12; Length 1313;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPETEGAPTSEE-----KKPIFGMKKPPGP 63
DB 365 PKPEPTTPPKPAPTKPAPTKPAPTKPAPTKPAPTKPAPTKPAP 409

RESULT 9
US-10-124-557-50

; Sequence 50, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 1314 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-124-557-50

Query Match 17.3% Score 76; DB 12; Length 1314;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPETEGAPTSEE-----KKPIFGMKKPPGP 63
DB 366 PKPEPTTPPKPAPTKPAPTKPAPTKPAPTKPAPTKPAPTKPAP 410

RESULT 10
US-10-124-557-46

; Sequence 46, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 17.3% Score 76; DB 12; Length 1320;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGOPRRKSTPETEGAPTSEE-----KKPIFGMKKPPGP 63
DB 372 PKPEPTTPPKPAPTKPAPTKPAPTKPAPTKPAPTKPAPTKPAP 416

RESULT 11
US-10-124-557-60

; Sequence 60, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Geert, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

LENGTH: 1320 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-124-557-46

Query Match 17.3% Score 76; DB 12; Length 1320;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

```

;
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
;
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
;
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
;
; US-10-124-557-60
;
; Query Match 17.3%; Score 76; DB 12; Length 1320;
; Best Local Similarity 42.2%; Pred. No. 15;
; Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
;
; QY 25 PGAGQPPRRKSTPETEGAPTSEE-----KKPIGMMKKFGP 63
; Db 372 PKPPTPTTKKPAPPTTKKPAPPTTKKPAPPTAKKKAPPTTKKPAP 416
;
; RESULT 12
; US-10-124-557-48
; Sequence 48, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
;
; US-10-124-557-48
```

```

;
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
;
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
;
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
;
; US-10-124-557-48
;
; Query Match 17.3%; Score 76; DB 12; Length 1354;
; Best Local Similarity 42.2%; Pred. No. 15;
; Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
;
; QY 25 PGAGQPPRRKSTPETEGAPTSEE-----KKPIGMMKKFGP 63
; Db 406 PKPPTPTTKKPAPPTTKKPAPPTTKKPAPPTAKKKAPPTTKKPAP 450
;
; RESULT 13
; US-10-124-557-40
; Sequence 40, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
;
; US-10-124-557-40
```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1361 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein.
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-124-557-40

Query Match 17.3%; Score 76; DB 12; Length 1361;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
US-10-124-557-2

QY 25 PGACQPPRRKSTPETEGAPTTSEE-----KKPIPGMKKPPGP 63
DB 413 PKEPTPTTPKEPAPTTKEPAPTTKEPAPTPAKKAPPTTPKEPAP 457

RESULT 14
US-10-124-557-52
Sequence 52, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gessner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann

REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 1363 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 17.3%; Score 76; DB 12; Length 1363;
Best Local Similarity 42.2%; Pred. No. 15;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
US-10-124-557-2

QY 25 PGACQPPRRKSTPETEGAPTTSEE-----KKPIPGMKKPPGP 63
DB 415 PKEPTPTTPKEPAPTTKEPAPTTKEPAPTPAKKAPPTTPKEPAP 459

RESULT 15
US-10-124-557-2
Sequence 2, Application US/10124557
Patent No. US20020137894A1
GENERAL INFORMATION:
APPLICANT: Turner, Katherine
Clark, Stephen C.
Jacobs, Kenneth
Hewick, Rodney M.
Gessner, Thomas G.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/124,557
FILING DATE: 16-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseerr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-124-557-2

Query Match 17.3%; Score 76; DB 12; Length 1404;
Best Local Similarity 42.2%; Pred. Nd. 16;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGAGQPPRRKSTPETEREGAPTSEE-----KKPIPGMKKFGP 63
DB 456 PKETPTTPKEPAPTTPKEPAPTTPKEPAPTAPKKKAPTTPEKPA 500

Search completed: April 9, 2003, 12:53:39
Job time : 22.8508 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:29:50 ; Search time 30.504 Seconds
(without alignments)
371.305 Million cell updates/sec

Title: US-09-647-019-2

Perfect score: 440
Sequence: 1 MSKQPSNVRALQANINIPM.....NLSEIQNVKSEKFKPKGQ 85

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	440	100.0	85	AA198650	Murine chisel (Csl)
2	382.5	86.9	88	AA198651	Human chisel (Csl)
3	382.5	86.9	88	AA198652	Human chisel (Csl)
4	262.5	59.7	75	AA198834	Xenopus chisel (Cs)
5	77	17.5	2153	AA193195	Novel human secret
6	76	17.3	1299	AA194322	Human EST encoded
7	76	17.3	1404	AA196049	MSF precursor. Sy
8	76	17.3	1404	AA196058	Human megakaryocyte
9	76	17.3	1404	AA196058	Human megakaryocyte
10	76	17.3	1415	AA196262	Novel human secret

11	75.5	17.2	771	22	AB198529
12	74	16.8	411	23	AB193428
13	73.5	16.7	91	22	AA004777
14	73	16.6	222	21	AA194997
15	72.5	16.5	249	22	AB197048
16	72	16.4	2518	21	AA1940574
17	71.5	16.2	143	22	AA1932495
18	71	16.1	1219	22	AB1965388
19	70.5	16.0	356	20	AA197714
20	70.5	16.0	356	23	AA1975492
21	70.5	16.0	363	18	AA1975492
22	70.5	16.0	521	22	AA197304
23	70.5	16.0	603	22	AA197594
24	70.5	16.0	603	22	AA197140
25	70.5	16.0	1908	22	AB1970137
26	70.5	16.0	2234	22	AB1970137
27	70	15.9	154	22	AA1975492
28	70	15.9	976	22	AA1975492
29	70	15.9	976	22	AA1975492
30	69.5	15.8	208	20	AA1975492
31	69	15.7	146	22	AA1975492
32	69	15.7	902	22	AA1975492
33	69	15.7	1001	22	AB1975492
34	69	15.7	1098	22	AB1975492
35	68.5	15.6	182	22	AB1975492
36	68.5	15.6	1076	22	AA1975492
37	68.5	15.6	1076	23	AA1975492
38	68.5	15.6	2061	22	AB1975492
39	68	15.5	499	22	AB1975492
40	68	15.5	499	22	AB1975492
41	68	15.5	1718	22	AB1975492
42	67.5	15.3	100	22	AA1975492
43	67.5	15.3	430	22	AA1975492
44	67.5	15.3	568	18	AA1975492
45	67	15.2	159	22	AA1975492

ALIGNMENTS

RESULT 1	AA198650	standard; Protein; 85 AA.
XX	AA198650;	
AC	17-JAN-2000	(first entry)
DT	17-JAN-2000	(first entry)
XX		
DE	Murine chisel (Csl) protein.	
XX		
KW	Chisel protein; Csl; EF-Hand protein super family; muscle development;	Drosophila melanog
KW	heart/skeletal muscle cell development; signalling pathway; murine;	Herbicidally activ
KW	X chromosome; regulation; adaptive process; muscle homeostasis; nucleus;	Human polypeptide
KW	detection; diagnosis; prophylaxis; treatment; skeletal myopathy;	Arabidopsis thalia
KW	muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;	Novel human diagn
KW	Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;	Novel human secret
KW	differentiation; gene therapy; transgenic animal; drug screening;	Drosophila melanog
KW	scallop regulatory myosin light chain.	Staphylococcus aur
XX		
OS	Mus sp.	Staphylococcus aur
XX		
PN	WO950410-A1.	Novel central nerv
XX		
PD	07-OCT-1999.	Novel signal trans
XX		
PF	26-MAR-1999;	Novel human diagn
XX	99WO-AU00220.	Novel human secret
XX		
PR	27-MAR-1998;	Staphylococcus aur
XX	98AU-0002634.	S. aureus antigeni
XX		
PA	(CHAN-) CHANG CARDIAC RES INST VICTOR.	Staphylococcus aur
PA	(GEHO-) GEN HOSPITAL CORP.	Novel central nerv
PA	(HALL-) HALL INST MEDICAL RES WALTER & ELIZA.	Novel signal trans
XX		

PI Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX WPI: 1999-610852/52.
 DR N-PSDB; AAX90903.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 2; Fig 3; 157pp; English.
 XX
 XX The present sequence is the murine chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 SQ Sequence 85 AA;
 Query Match Best Local Similarity 100.0%; Score 440; DB 20; Length 85;
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKQPISNVRAIQININIPMGAFRPGAGOPRRKSTPETEBGAPTTSEKKPIPGMKKF 60
 DB 1 MSKQPISNVRAIQININIPMGAFRPGAGOPRRKSTPETEBGAPTTSEKKPIPGMKKF 60
 QY 61 PGPVNLSEIQNVKSELKFPKGEQ 85
 DB 61 PGPVNLSEIQNVKSELKFPKGEQ 85
 ID AAY28651
 ID AAY28651 standard; Protein: 88 AA.
 XX
 AC AAY28651;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Human chisel (Csl) protein.
 XX
 KW Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; regulation;
 KW XG21.3-q22; adaptive process; muscle homeostasis; cardiac hypertrophy;
 KW detection; diagnosis; prophylaxis; treatment; differentiation; nucleus;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW skeletal myopathy; Becker's myotonic dystrophy; heart failure;
 KW transgenic animal; drug screening; gene therapy; homology;
 KW scallop regulatory myosin light chain.
 XX
 OS Homo sapiens.
 XX
 PN WO950410-A1.
 XX
 PD 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WO-AU00220.
 PF
 XX 27-MAR-1998; 98AU-0002634.
 PR
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA

PA (GEO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PT Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX WPI: 1999-610852/52.
 DR N-PSDB; AAX90904.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 7; Fig 3; 157pp; English.
 XX
 XX The present sequence is the human chisel protein (Csl), that is a member
 CC of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and has 86% homology
 CC to the mouse Csl. Structural homology between Csl and scallop regulatory
 CC myosin light chain is also detected. It is predominantly expressed in
 CC heart and skeletal muscles and is activated after the differentiation of
 CC cells. Csl functions in regulation aspects of differentiation or
 CC adaptive processes that maintain muscle homeostasis. This sequence can
 CC be used in the detection, diagnosis, prophylactic and therapeutic
 CC treatment of diseases such as those involving aberrant muscle cell
 CC development and functional activity. It is also used in the treatment of
 CC muscular and myotonic dystrophies, skeletal myopathies such as Duchenne
 CC muscular dystrophy and Becker's myotonic dystrophy, heart failure,
 CC cardiac hypertrophy, myocarditis, myofiber atrophy, etc. The Csl gene
 CC sequence can also be used in gene therapy, for the production of
 CC transgenic animals and for drug screening.
 XX
 SQ Sequence 88 AA;
 Query Match Best Local Similarity 86.3%; Score 382.5; DB 20; Length 88;
 Matches 74; Conservative 3; Mismatches 8; Indels 1; Gaps 1;
 QY 1 MSKQPISNVRAIQININIPMGAFRPGAGOPRRKSTPETEBGAPTTSEKKPIPGMKK 59
 DB 3 MSKQPISNVRAIQININIPMGAFRPGAGOPRRKSTPETEBGAPTTSEKKPIPGAKK 62
 QY 60 PGPVNLSEIQNVKSELKFPKGEQ 85
 DB 63 LRGPAVNLSEIQNVKSELKFPKGEQ 88
 ID AAE16632
 ID AAE16632 standard; Protein: 88 AA.
 XX
 AC AAE16632;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human 66214 protein.
 XX
 KW Human; congestive heart failure; dilative cardiomyopathy; sudden death;
 KW hypertrophic cardiomyopathy; ischemic cardiomyopathy; rhythm disorder;
 KW heart muscle disease; conduction disorder; coronary heart disease;
 KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
 KW pulmonary heart disease; valvular heart disease; pericardial disease;
 KW congenital heart disease; gene therapy; syncope; transgenic animal;
 KW expressed sequence tag; EST; 66214 protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200192567-A2.
 XX
 PD 06-DEC-2001.
 XX
 XX 30-MAY-2001; 2001WO-EP06165.
 PF
 XX 30-MAY-2000; 2000US-207400P.
 PR

XX (MED1-) MEDIGENE AG.
 XX Bunk D, Reuner B, Beck J, Henkel T;
 XX WPI: 2002-123073/16.
 DR N-PSDB; AAD27216.
 XX
 PT Identifying a subject at risk for a heart disease e.g. congestive heart
 PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 PT the polypeptide expressed by genes abnormally expressed in heart tissue
 PT
 XX Claim 1a; Fig 9c; 154pp; English.
 PS
 XX The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present sequence is 66214 protein encoded by an
 CC expressed sequence tag (EST) DNA.
 XX
 SQ Sequence 88 AA;
 Query Match 86.9%; Score 382.5; DB 23; Length 88;
 Best Local Similarity 86.0%; Pred. No. 6e-34;
 Matches 74; Conservative 3; Mismatches 8; Indels 1; Gaps 1;
 QY 1 MSKQPSNVRAIQANINIPMGAFRPGAGOPRRKKESTPETEGAPTTS-BEKKPIPGMKK 59
 DB 3 MSKQPSNVRAIQANINIPMGAFRPGAGOPRRKKECTPEVEGVPPISDEBKRTIPAKK 62
 QY 60 PRGVNLTSEIQNKSEIKFVPKGEQ 85
 DB 63 LPGAVALSEIQNKSEIKFVPKAEQ 88
 RESULT 4
 AAY28834
 ID AAY28834 standard; Protein; 75 AA.
 XX
 AC AAY28834;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Xenopus chisel (Csl) protein.
 XX
 XX Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KM heart/skeletal muscle cell development; signalling pathway; xenopus;
 KM X chromosome; regulation; adaptive process; muscle homeostasis; nucleus;
 KM detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KM Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KM differentiation; gene therapy; transgenic animal; drug screening;
 KM scallop regulatory myosin light chain.
 XX
 OS Xenopus sp.
 XX
 PN MO950410-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 26-MAR-1999; 99MO-AU00220.

XX
 PR 27-MAR-1998; 98AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PI Harvey RP, Musaro A, Palmer SJ, Rosenchal NA;
 XX
 DR WPI: 1999-610852/52.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 11; Fig 3; 157pp; English.
 XX
 XX The present sequence is the xenopus chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 SQ Sequence 75 AA;
 Query Match 59.7%; Score 262.5; DB 20; Length 75;
 Best Local Similarity 70.3%; Pred. No. 4.9e-21;
 Matches 52; Conservative 5; Mismatches 12; Indels 5; Gaps 1;
 QY 1 MSKQPSNVRAIQANINIPMGAFRPGAGOPRRKKESTPETEGAPT-----TSEKKPIIP 55
 DB 1 MSKQPSNVRAIQANINIPMGAFRPGAGOPRRKKESTPEEGHVPPESEKSEKKPIIP 60
 QY 56 GMKKPPGPVNLTSE 69
 DB 61 GAVKLPGPAPNLTSE 74
 RESULT 5
 AAU33195
 ID AAU33195 standard; Protein; 2153 AA.
 XX
 AC AAU33195;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #3686.
 XX
 XX Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001MO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 XX
 PR 26-JAN-2001; 2001US-0770160.

```

XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX XX WPI; 2001-611725/70.
XX DR
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20; Page 727-728; 765pp; English.
XX XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to, a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation, to regulate haematopoiesis, and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemia. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.
XX XX
SQ Sequence 2153 AA;
Query Match 17.5%; Score 77; DB 22; Length 2153;
Best Local Similarity 30.4%; Pred. No. 28;
Matches 24; Conservative 10; Mismatches 31; Indels 14; Gaps 3;
QY 3 KQPTSNRAIQANINIPWAGPRGQPPRRKSPETEEGAPTTSEKKRP----- 55
DB 875 KEPTAQLTGFPAGKRLP--ASRSAGAPPPRRSEKTRTWG-PPGSSRGILPPEEGCAP 931
QY 56 ----GMKKFPGPVVNLSEI 70
DB 932 PRRAGPIKKPPPTKVEEL 950
RESULT 6
AAM24322
ID AAM24322 standard; Protein: 1299 AA.
XX AC AAM24322;
XX DT 12-OCT-2001 (first entry)
XX DE Human EST encoded protein SEQ ID NO: 1847.
XX XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX OS Homo sapiens.
XX PN W0200154477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02687.
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX FT

```

```

PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Werthman T;
XX XX WPI; 2001-476164/51.
XX DR N-PSDB; AAH98981.
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX PT antibodies and research use -
XX PS Claim 20; Page 1198-1201; 1275pp; English.
XX XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC for analysis, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX XX
SQ Sequence 1299 AA;
Query Match 17.3%; Score 76; DB 22; Length 1299;
Best Local Similarity 42.2%; Pred. No. 20;
Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
QY 25 PGAGQPPRRKSPETEEGAPTTSE-----KKPIPGMKKPPGP 63
DB 456 PKETPTTPKPEAPTTKEPAPTTKPEAPTPKAPKPPAPTTKPEAP 500
RESULT 7
AAR26049
ID AAR26049 standard; Protein: 1404 AA.
XX AC AAR26049;
XX DT 02-FEB-1993 (first entry)
XX DE MSF precursor.
XX XX
KW Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;
KW stability; proteolytic cleavage; adhesion; alternative splicing.
XX OS Synthetic.
XX FH Key
XX FT Location/Qualifiers
XX FT 1..26
XX FT /label= Exon_I
XX FT 26..67
XX FT /label= Exon_II
XX FT 67..107
XX FT /label= Exon_III
XX FT 107..157
XX FT /label= Exon_IV
XX FT 157..200
XX FT /label= Exon_V
XX FT 200..1141
XX FT /label= Exon_VI
XX FT 1411..1166
XX FT /label= Exon_VII
XX FT 1166..1212
XX FT /label= Exon_VIII
XX FT 1213..1266
XX FT /label= Exon_IX
XX FT 1266..1331
XX FT /label= Exon_X
XX FT 1331..1373
XX FT /label= Exon_XI
XX FT 1373..1404
XX FT /label= Exon_XII

```


XX WO9213075-A.
 XX 06-AUG-1992.
 XX 17-JUN-1992; 92MO-US00433.
 XX 18-JAN-1991; 91US-0643502.
 XX 10-SEP-1991; 91US-0757022.
 XX (GENY) GENETICS INST INC.
 XX
 XX Clark SC, Geener TG, Hewick RM, Jacobs K, Turner K;
 XX WPI; 1992-284660/34.
 XX N-PSDB; AAQ27223.
 XX
 XX New human mega-karyocyte stimulating factors - for treating
 XX immune deficiencies, cancer, exposure to radiation or drugs,
 XX bacterial and viral infections, etc.
 XX
 XX Claim 1, 2 and 3; Fig 1; 87pp; English.

XX The sequence given is a full length translation from the megakaryocyte
 XX stimulating factor (MSF) precursor. The sequence covered by exons II,
 XX III and IV encodes megakaryocyte stimulating factor (MSF). This
 XX sequence is modified by the addition of an N-terminal sequence encoding
 XX a secretory leader, an initiating methionine proceeding exon II and a
 XX terminating codon following exon IV. The cDNA sequence given contains
 XX (meg-CSF). Exon I contains the initiating methionine, and encodes a
 XX classical mammalian protein secretion signal sequence. The sequence
 XX encoding the original meg-CSF includes exons II-IV and is thought to
 XX terminate in the region between amino acid residues 134 - 147. The
 XX primary transcript of this gene may be cleaved in different ways to
 XX yield a family of mRNAs each encoding a different MSF protein. Exons
 XX V and VI are thought to be related to the activity of the factor and
 XX are also implicated in the stability, folding and processing of the
 XX molecule. These exons are also thought to play a role in the observed
 XX synergy of MSF with other cytokines. Exons V - XII are believed to be
 XX implicated in the processing or folding of the appropriate structure of
 XX the resulting factor. ie. one or more of these exons may contain
 XX sequences which direct proteolytic cleavage, adhesion, organisation of
 XX the cellular matrix or extracellular matrix processing. Both naturally
 XX occurring and non-naturally occurring MSF's may be characterised by
 XX various combinations of alternatively spliced exons from this sequence,
 XX with the exons spliced together in differing orders to form different
 XX members of the MSF family.

XX Sequence 1404 AA;

Query Match 17.3%; Score 76; DB 13; Length 1404;
 Best Local Similarity 42.2%; Pred. No. 22;
 Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGACQPPRRKSTPTEBGAFTTSEB-----KKPIPMKKKPPGP 63
 DB 456 PKBPTPTPKBPAPPTTKBPAPPTPKBPAPPKKBPAPPTPKBPAP 500

RESULT 8
 AAB60568
 ID AAB60568 standard; Protein; 1404 AA.
 XX
 XX AAB60568;
 AC
 XX
 XX 27-APR-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF, CACP).
 DE Human megakaryocyte stimulating factor (MSF, CACP).
 XX
 XX Human; CACP protein; campodactylly-arthropathy-coxa vara-pericarditis;
 KW MSF; megakaryocyte stimulating factor; synovial lubricant;
 KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;

KM antiarthritic.
 XX
 XX Homo sapiens.
 OS
 XX
 XX MO200107068-A1.
 XX
 XX 01-FEB-2001.
 XX
 XX 21-JUL-2000; 2000MO-US20002.
 XX
 XX 23-JUL-1999; 99US-0145328.
 XX 19-JUL-2000; 2000US-0145328.
 XX
 XX (UYCA-) UNIV CASE WESTERN RESERVE.
 XX
 XX Warman ML;
 XX WPI; 2001-182721/18.
 XX
 XX New composition comprising the campodactylly-arthropathy-coxa
 XX vara-pericarditis protein in combination with an anesthetic, useful for
 XX treating osteoarthritis, or as lubricants of tissue and joints -
 XX
 XX Example 1; Page -; 34pp; English.

XX The invention relates to a method of treating osteoarthritis via the
 XX administration of a composition comprising the campodactylly-arthropathy-
 XX coxa vara-pericarditis (CACP) protein, or portions of the CACP protein.
 XX The composition may further comprise a local anesthetic. The composition
 XX of the invention may be administered via intra-articular or intravenous
 XX injection. The human CACP protein is identified in the invention as
 XX being megakaryocyte stimulating factor (MSF). The gene encoding
 XX CACP protein (MSF) is located on chromosome 1q25-31, and mutations in
 XX this gene are responsible for the heritable disorder campodactylly-
 XX arthropathy-coxa vara-pericarditis, in which patients have synovial
 XX hyperplasia without evidence of inflammation. CACP protein (MSF)
 XX acts as a synovial lubricant, and can be used to lubricate tissue and
 XX joints in the treatment of osteoarthritis. The composition may be
 XX applied to reduce the symptoms of osteoarthritis (e.g., joint pain,
 XX loss of range of movement or joint damage). The present sequence
 XX represents human megakaryocyte stimulating factor (MSF, CACP protein).
 XX Note: This sequence is not given in its entirety in figure 4 of the
 XX specification, although a GenBank accession number was given. This
 XX sequence was therefore obtained from GenBank (U70316).

XX Sequence 1404 AA;

Query Match 17.3%; Score 76; DB 22; Length 1404;
 Best Local Similarity 42.2%; Pred. No. 22;
 Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;

QY 25 PGACQPPRRKSTPTEBGAFTTSEB-----KKPIPMKKKPPGP 63
 DB 456 PKBPTPTPKBPAPPTTKBPAPPTPKBPAPPKKBPAPPTPKBPAP 500

RESULT 9
 AAB29773
 ID AAB29773 standard; Protein; 1404 AA.
 XX
 XX AAB29773;
 AC
 XX
 XX 28-FEB-2001 (first entry)

XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.
 DE Human megakaryocyte stimulating factor; tribonectin;
 XX
 XX Human MSF; megakaryocyte stimulating factor; tribonectin;
 KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
 KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
 KW friction coefficient reduction; gene therapy; antiarthritic;
 KW osteopathic.
 XX
 XX Homo sapiens.
 OS

```

XX PN WO200064930-A2.
XX PD 02-NOV-2000.
XX PF 24-APR-2000; 2000WO-US10953.
XX PR 23-APR-1999; 99US-0298970.
XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX PI Jay GD;
XX DR WPI; 2001-024673/03.
XX DR N-PSDB; AAC81498.
XX PT Novel tribonectin polypeptide useful as lubricant for treating
XX PT osteoarthritis, comprises O-linked lubricating moiety -
XX PS Claim 3, Page 7, 47pp; English.
XX CC The invention relates to a human tribonectin which is a product of
XX CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
XX CC gene. The tribonectin has at least one O-linked oligosaccharide
XX CC lubricating moiety and has a polypeptide sequence comprising 1-76
XX CC repeats of a motif having at least 50% identity to the sequence KKPAPTT
XX CC (AAB29774). The invention also relates to a nucleic acid encoding a
XX CC human MSF-derived tribonectin; a biocompatible composition comprising a
XX CC human tribonectin for inhibiting tissue adhesion formation; and a method
XX CC of diagnosing osteoarthritis or a predisposition to osteoarthritis by
XX CC measuring the amount of MSF or its fragment in a biological sample of a
XX CC mammal, wherein an increased amount of MSF compared to a control
XX CC indicates the presence of or predisposition to developing
XX CC osteoarthritis. The tribonectin and DNA encoding it are useful in the
XX CC treatment of osteoarthritis, where they may be used for lubricating
XX CC mammalian joints, such as articulating joints of humans, dogs or horses.
XX CC The tribonectin, when formulated as a membrane, foam, gel or fibre, is
XX CC useful for inhibiting adhesion between two surfaces such as the injured
XX CC tissues of a mammal, where the injury is caused by a surgical insertion,
XX CC or trauma, or an artificial device e.g., an orthopaedic implant. In
XX CC particular, one of the surfaces is pericardial tissue. DNA encoding a
XX CC tribonectin may be used in gene therapy. The present sequence represents
XX CC human MSF.
XX CC
XX SO Sequence 1404 AA;
XX
XX Query Match 17.3%; Score 76; DB 22; Length 1404;
XX Best Local Similarity 42.2%; Pred. No. 22;
XX Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
XX
XX QY 25 PGACGPPRRKSTPTEGAPTTSE-----KKPIPKMKKPPGP 63
XX DB 456 PKEPPTTPKPEAPTTKPEAPTTKPEAPTPAPKPKAPPTTKPEAP 500
XX
XX RESULT 10
XX ID AU32262
XX ID AU32262 standard; Protein; 1415 AA.
XX AC AAU32262;
XX AC
XX DT 18-DEC-2001 (first entry)
XX DB Novel human secreted protein #2753.
XX KW Human; vaccination; gene therapy; nutritional supplement;
XX KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
XX KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX OS Homo sapiens.
XX OS WO200179449-A2.
XX PN
XX

```

```

PD PD 25-OCT-2001.
XX PF 16-APR-2001; 2001WO-US08656.
XX PF 18-APR-2000; 2000US-0552929.
XX PR 26-APR-2001; 2001US-0770160.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-611725/70.
XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic
XX PT vaccination, testing and therapy -
XX PS Claim 20; Page 573; 765pp; English.
XX CC The invention relates to novel human secreted polypeptides. The
XX CC polypeptides and antibodies to the polypeptides are useful for
XX CC determining the presence of or predisposition to a disease associated
XX CC with altered levels of polypeptide. The polypeptides are also useful for
XX CC identifying agents (agonists and antagonists) that bind to them. Cells
XX CC expressing the proteins are useful for identifying a therapeutic agent
XX CC for use in treatment of a pathology related to aberrant expression or
XX CC physiological interactions of the polypeptide. Vectors comprising
XX CC the nucleic acids encoding the polypeptides and cells genetically
XX CC engineered to express them are also useful for producing the proteins.
XX CC The proteins are useful in genetic vaccination, testing and
XX CC therapy, and can be used as nutritional supplements. They may be used to
XX CC increase stem cell proliferation; to regulate haematopoiesis; and in
XX CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
XX CC immune suppression and/or stimulation; as anti-inflammatory agents; and
XX CC in treatment of leukaemias. AA029510-AA03304 represent the amino acid
XX CC sequences of novel human secreted proteins of the invention.
XX CC
XX SO Sequence 1415 AA;
XX
XX Query Match 17.3%; Score 76; DB 22; Length 1415;
XX Best Local Similarity 42.2%; Pred. No. 22;
XX Matches 19; Conservative 2; Mismatches 18; Indels 6; Gaps 1;
XX
XX QY 25 PGACGPPRRKSTPTEGAPTTSE-----KKPIPKMKKPPGP 63
XX DB 467 PKEPPTTPKPEAPTTKPEAPTTKPEAPTPAPKPKAPPTTKPEAP 511
XX
XX RESULT 11
XX ID ABB58529
XX ID ABB58529 standard; Protein; 771 AA.
XX AC ABB58529;
XX AC
XX DT 26-MAR-2002 (first entry)
XX DB Drosophila melanogaster polypeptide SEQ ID NO 2379.
XX DE Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX KW Drosophila melanogaster.
XX OS Drosophila melanogaster.
XX OS WO200171042-A2.
XX PN
XX PD 27-SEP-2001.
XX PD
XX PF 23-MAR-2001; 2001WO-US09231.
XX PF 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX

```

PI Venter JC, Adams M, Li PWD, Myers BW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL02632.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 2379; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL1840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 771 AA;
 XX
 Query Match 17.2%; Score 75.5; DB 22; Length 771;
 Best Local Similarity 30.6%; Pred. No. 12;
 Matches 15; Conservative 11; Mismatches 14; Indels 9; Gaps 2;
 XX
 QY 25 PGACGPPRRKSTPTEGAPTSTSEKK-----PIP-GMKKPGPV 64
 DB 268 PRLGQPPQQQQQQQQPQQQTPTQQQAQSQAAASLPLVPHGMPGMPGM 316
 XX
 RESULT 12
 ABB93428
 ID ABB93428 standard; Protein; 411 AA.
 XX
 AC ABB93428;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 2639.
 XX
 KM Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (PARB) BAYER AG.
 XX
 PI Tietjen K, Weidner M;
 XX
 DR WPI; 2002-269010/31.
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 2639; 26pp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins
 CC (ABB9790-ABB9406) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 XX
 SQ Sequence 411 AA;
 XX
 Query Match 16.8%; Score 74; DB 23; Length 411;
 Best Local Similarity 33.7%; Pred. No. 8.8;
 Matches 29; Conservative 9; Mismatches 24; Indels 24; Gaps 5;
 XX
 QY 2 SKQP-ISNVRATQANINIPMGAFRPGAGPPRRKES-----TPETEGAPT-----46
 DB 86 SKQLFISYVESYARFNI-----KPVNQTVKAEAFDDASGLMNVKIQGVSTWLVVA 140
 XX
 QY 47 TSEKKP-----IPGMKPPGPPVNL 68
 DB 141 TGENAPVFPFNIPLKKFTGPPVHTS 166
 XX
 RESULT 13
 AA004777
 ID AA004777 standard; Protein; 91 AA.
 XX
 AC AA004777;
 XX
 DB 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 18669.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 XX
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI84708.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 18669; 139pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

```
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 91 AA;
Query Match 16.7%; Score 73.5; DB 22; Length 91;
Best Local Similarity 34.4%; Pred. No. 1.8;
Matches 21; Conservative 7; Mismatches 14; Indels 19; Gaps 4;

Cy 19 PMG-AFRGAGQ---PPRRKSTPTBEGAPTTSEKKPIPG-----MKKTPG 62
Db 27 PIGKTRMGKKEKPPPPGKK---KKGGPPTLGKKNPVPQMGAFLLPLEKKKKPPFG 83

Cy 63 P 63
Db 84 P 84

RESULT 14
AAG49497
ID AAG49497 standard; Protein; 222 AA.
XX
AC AAG49497;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SRQ ID NO: 62625.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127442.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130831.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 16-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
```

```
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
```

```

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149922.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 16.4%; Score 73; DB 21; Length 222;
Best Local Similarity 33.9%; Pred. No. 5, 6;
Matches 21; Conservative 8; Mismatches 21; Indels 12; Gaps 2;

QY 13 QANINIPMGAFRPGAGPPRRKSTPTETEGAPPTTSEKRPROMKFPDPVNLSTQN 72
DB 58 QANRLLNGSFPGSGMIPRKESWPS---STTTDE-----FEKLMATPDEKN 105
QY 73 VK 74
DB 106 TK 107

RESULT 15
ABG27048
ID ABG27048 standard; Protein; 249 AA.
XX
AG ABG27048;
XX
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #27039.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
PR 23-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS91235.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 57407; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

```


C	7	560	63.1	683	12	BF111459	BF111459 7134h12.x
	8	554.2	62.5	306	12	BF672126	BF672126 602152580
	9	538	60.7	879	12	BF575112	BF575112 602134792
C	10	531.8	60.0	587	9	AA211443	AA211443 zms5b01.8
	11	518	58.4	541	14	BM717052	BM717052 UI-E-EJO-
C	12	518	58.4	544	14	BM674432	BM674432 UI-E-EJO-
C	13	495	55.8	495	9	AI090520	AI090520 qa70d01.x
C	14	493	55.6	494	10	AM170011	AM170011 xj35g08.x
	15	469.6	52.9	501	9	AA211521	AA211521 zms5b01.r
	16	457.8	51.6	934	11	AK010172	AK010172 MUB mucu
	17	456.2	51.4	871	12	BF693124	BF693124 602080151
C	18	452	51.0	452	9	AI190689	AI190689 qd50d01.x
	19	450.2	50.8	614	12	BG223758	BG223758 IM00018G0
C	20	449	50.6	455	9	AI355676	AI355676 qe56e08.x
	21	443.8	50.0	934	11	AK003105	AK003105 MUB mucu
C	22	438	49.4	439	9	AI333485	AI333485 qd88f03.x
	23	428.8	48.3	890	9	AV088480	AV088480 AV088480
	24	425.6	48.0	582	12	BG224246	BG224246 IM0030E01
	25	397.4	44.8	544	10	BE014221	BE014221 125832 MA
	26	396.2	44.7	826	12	BF126275	BF126275 601650456
	27	391.8	44.2	528	12	BF073866	BF073866 220906 MA
	28	389.2	43.9	517	12	BG224245	BG224245 IM0028F05
	29	383	43.2	521	12	BF261282	BF261282 IM0004B05
	30	383	43.2	867	12	BF790636	BF790636 602250354
	31	363.2	40.9	368	14	F36800	F36800 HSPD34697 H
	32	360	40.6	360	9	AI355905	AI355905 3055-10R-
	33	356.8	40.2	394	9	AA389647	AA389647 M04 Ret
	34	342.6	38.6	499	12	BG224307	BG224307 IM0033A12
C	35	330	37.2	330	9	AI355903	AI355903 BH11-PI-C
	36	330	37.2	330	9	AI355904	AI355904 BH11-RI-C
	37	319	36.0	466	12	BG358790	BG358790 BOWMS1-01
	38	309.8	34.9	313	14	F26916	F26916 HSPD14595 H
	39	308.2	34.7	480	9	AA214155	AA214155 zms5b01.r
	40	307	34.6	320	14	M07478	M07478 zms5b01.r
	41	305.2	34.4	412	12	BG223652	BG223652 IM00010F0
	42	302.8	33.9	426	10	BE013988	BE013988 125544 MA
C	43	300.8	33.9	321	9	AI824667	AI824667 wc48c01.x
	44	299.2	33.7	613	14	BO554133	BO554133 H4026C08-
	45	296.4	33.4	486	9	AI035961	AI035961 ud50b12.r

ALIGNMENTS

RESULT 1
 BM697544
 LOCUS
 DEFINITION UI-E-DXO-agn-1-12-0-UI.r1 UI-E-DXO Homo sapiens CDNA clone
 ACCESSION BM697544
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 719)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 GenBank: F06091
 JOURNAL MEDLINE 97044477
 COMMENT
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Ricketts Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 18:42:01 / Search time 52.4832 Seconds

(without alignments)
5183.040 Million cell updates/sec

Title: US-09-647-019-3

Perfect score: 887
1 ggtctccatcacccggagagag.....tgactaatcgttaaatcgtga 887

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	836.4	94.3	909	US-09-484-970B-111	Sequence 111, App
2	60	6.8	7218	US-08-232-463-14	Sequence 14, App
3	39.4	4.4	14672	US-08-961-527-111	Sequence 111, App
4	38.8	4.4	1755	US-09-134-001C-1047	Sequence 1047, App
5	37.2	4.2	25002	US-08-961-527-48	Sequence 48, App
6	37.2	4.2	246240	US-08-724-394A-20	Sequence 20, App
7	37.2	4.2	246240	US-08-724-394A-21	Sequence 21, App
8	37.2	4.2	246240	US-08-724-394A-22	Sequence 22, App
9	35.4	4.0	3182	US-08-971-395-1	Sequence 1, App
10	35.4	4.0	3183	US-08-971-395-1	Sequence 1, App
11	35.2	3.9	569	US-09-328-111-766	Sequence 766, App
12	35	3.9	1383	US-08-630-822A-82	Sequence 82, App
13	35	3.9	1383	US-09-005-069-82	Sequence 82, App
14	35	3.9	1383	US-09-171-156A-28	Sequence 28, App
15	35	3.9	1704	US-08-630-822A-67	Sequence 67, App
16	35	3.9	1704	US-09-005-069-67	Sequence 67, App
17	35	3.9	1704	US-09-171-156A-26	Sequence 26, App
18	35	3.9	3982	US-08-947-823-4	Sequence 4, App
19	35	3.9	51952	US-08-947-823-1	Sequence 1, App
20	34.8	3.9	19124	US-08-487-828B-13	Sequence 13, App
21	34.6	3.9	3204	US-09-336-643A-26	Sequence 26, App
22	34	3.8	8302	US-09-234-827B-1	Sequence 1, App
23	33.8	3.8	6243	US-09-056-075-1	Sequence 1, App
24	33.4	3.8	1924	US-09-071-035-455	Sequence 455, App
25	33.4	3.8	3963	US-09-071-035-449	Sequence 449, App
26	33.4	3.8	3963	US-09-071-035-453	Sequence 453, App
27	33.2	3.7	860	US-08-998-416-287	Sequence 287, App

28	33.2	3.7	969	US-09-502-600-30	Sequence 30, App
29	33.2	3.7	986	US-08-557-146-1	Sequence 1, App
30	33.2	3.7	986	US-09-154-344-1	Sequence 1, App
31	33.2	3.7	1089	US-08-930-188-1	Sequence 1, App
32	33.2	3.7	1089	US-08-930-188-3	Sequence 3, App
33	33.2	3.7	1089	PCT-US96-04294-1	Sequence 1, App
34	33.2	3.7	1089	PCT-US96-04294-3	Sequence 3, App
35	33.2	3.7	8012	US-09-182-117-1	Sequence 1, App
36	33.2	3.7	8418	US-09-182-117-5	Sequence 5, App
37	33.2	3.7	8798	US-09-182-117-4	Sequence 4, App
38	33.2	3.7	10846	US-09-098-219B-5	Sequence 5, App
39	33.2	3.7	10900	US-09-098-219B-6	Sequence 6, App
40	33.2	3.7	13158	US-08-687-080-105	Sequence 105, App
41	32.8	3.7	1121	5256558-7	Sequence 7, App
42	32.8	3.7	2097	US-09-134-001C-795	Sequence 795, App
43	32.8	3.7	2973	US-09-133-962A-15	Sequence 15, App
44	32.8	3.7	4304	US-08-368-776A-1	Sequence 1, App
45	32.8	3.7	4304	PCT-US96-00419-1	Sequence 1, App

ALIGNMENTS

RESULT 1					
US-09-484-970B-111					
Sequence 111, Application US/09484970B					
Patent No. 6426186					
GENERAL INFORMATION:					
APPLICANT: Jones, Karen A.					
APPLICANT: Volkmuth, Wayne					
APPLICANT: Walker, Michael G.					
TITLE OF INVENTION: BONE REMODELING GENES					
FILE REFERENCE: PB-0014 US					
CURRENT APPLICATION NUMBER: US/09/484,970B					
CURRENT FILING DATE: 2000-01-18					
NUMBER OF SEQ ID NOS: 172					
SOFTWARE: PERL Program					
SEQ ID NO 111					
LENGTH: 909					
TYPE: DNA					
ORGANISM: Homo sapiens					
FEATURE:					
NAME/KEY: misc_feature					
OTHER INFORMATION: Incyte ID No. 6426186 021656.2CBI					
US-09-484-970B-111					
Query Match					
Best Local Similarity 94.3%; Score 836.4; DB 4; Length 909;					
Matches 870; Conservative 0; Mismatches 1; Indels 3; Gaps 3;					
2 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAAAGATCGGAATTGA 61					
36 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAAAGATCGGAATTGA 95					
62 GATGCACTGACGAGGACACCC-GGGCGCCCTTCCACCTTCCAAAGAGCTTTATTCTT 120					
96 GATGCACTGACGAGGACACCC-GGGCGCCCTTCCACCTTCCAAAGAGCTTTATTCTT 155					
121 GCATCT-GGCTGCTGAGGACTTCCCTTAGGACAGTAAACAAATCATTAACAGGATTAAG 179					
156 GCATCTGGGCTGAGGACTTCCCTTAGGACAGTAAACAAATCATTAACAGGATTAAG 215					
180 ACTGCATG-AATATGCGAAACAGCAGTTTCCAAATGTAAGCATCCAGGAAATATC 238					
216 ACTGCATGAAAAATGTGAAACAGCAGTTTCCAAATGTAAGCATCCAGGAAATATC 275					
239 AATATTCATGGAAGCTTTGCGCCAGAGCAGTCAACCCCGAGAAAGAAATGT 298					
276 AATATTCATGGAAGCTTTGCGCCAGAGCAGTCAACCCCGAGAAAGAAATGT 335					
299 ACTGCTGAAGTGAAGGAGGAGTCTCCACCTCGGATGAGGAGAAAGCAATTCCA 358					
336 ACTCTGAAGTGAAGGAGGAGTCTCTCCACCTCGGATGAGGAGAAAGCAATTCCA 395					

QY 359 GGAGCGAAGAACTTCAGAGACCTGAGTCATCTATCGGAAATCCAGATATTAAGT 418
 DB 356 GGAGCGAAGAACTTCAGAGACCTGAGTCATCTATCGGAAATCCAGATATTAAGT 455
 QY 419 GAAGTAAATATGTCCCAAGCTGAAACAGTAGAGAGAGAGAGAGAGAGAGAGAG 478
 DB 456 GAAGTAAATATGTCCCAAGCTGAAACAGTAGAGAGAGAGAGAGAGAGAGAGAG 515
 QY 479 AATTAAGGAG 538
 DB 516 AATTAAGGAG 575
 QY 539 ATTGTGAACCTCTGAAATGCTGAGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
 DB 576 ATTGTGAACCTCTGAAATGCTGAGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 635
 QY 599 CT 658
 DB 636 CT 695
 QY 659 AATTAAGGAG 718
 DB 696 AATTAAGGAG 755
 QY 719 GGAGAGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 778
 DB 756 GGAGAGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 815
 QY 779 AAGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 838
 DB 816 AAGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 875
 QY 839 TAACTTGTATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 872
 DB 876 TAACTTGTATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 909

RESULT 2

US-08-232-463-14/C
 Sequence 14, Application US/08232463
 Patent No. 5670367

GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEIFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-9300
 TELEFAX: (703) 683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: pTZgpc-Fls
 US-08-232-463-14

Query Match 6.8%; Score 60; DB 1; Length 7218;
 Best Local Similarity 2.4%; Pred. No. 2.6e-07;
 Matches 9; Conservative 222; Mismatches 137; Indels 0; Gaps 0;

QY 147 AGGAGTAAACAATACATAAGCAGAGATGAGCTGATGATATGCGAAGCCAG 206
 DB 1420 RRR 1361
 QY 207 TTTCATGTGAGAGCCATCCAGCAATATCAATATTCATGAGAGAGCTTCGGCCAG 266
 DB 1360 RRR 1301
 QY 267 GAGAGGTCAACCCCGAGAGAGAGAGAGATGTACTCTGAGAGAGAGAGAGAGAGAG 326
 DB 1300 RRR 1241
 QY 327 CCACCTGGATGAG 386
 DB 1240 RRR 1181
 QY 387 TCATCTATGAGAGATCCAGATATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 446
 DB 1180 RRR 1121
 QY 447 AGTAGTAG 506
 DB 1120 RRR 1061
 QY 507 AGCTCACT 514
 DB 1060 AGCTCCT 1053

RESULT 3

US-08-961-527-111/C
 Sequence 111, Application US/08961527
 Patent No. 6420135

GENERAL INFORMATION:
 APPLICANT: Charles Kunesh
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,527
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 14672 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-111

Query Match 4.4%; Score 39.4; DB 4; Length 14672;
Best Local Similarity 51.4%; Pred. No. 0.24;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 707 ATATGATTTATGAGATGATGAGCAATGCTACAGATTGATGAAAGTTCCC 766
DB 562 ATATATCTGCTTAAATATACATTAGTGAATGAAAGCAATATATATATAGT 503
QY 767 AAGTACTCTACAGATATTTGTCATATTTGGAATGCGTTTACCTTTT 826
DB 502 AATTAACATCTACAGCAATGCAATGCGTCTGAAATGCGTCAATATATATAT 443
QY 827 AATTAATGCTAACTTTGATGATGATGCAATATATTTGCTAAATATAT 883
DB 442 AGCTATAGTACTAATGATGATGATGATGATGATGATGATGATGATGAT 386

RESULT 4
US-09-134-001C-1047
Sequence 1047, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1047
LENGTH: 1755
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1047

Query Match 4.4%; Score 38.8; DB 4; Length 1755;
Best Local Similarity 46.1%; Pred. No. 0.15;
Matches 130; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 523 ATATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
DB 1373 AAACAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1432
QY 583 GTTGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 642
DB 1433 TACTTCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1492
QY 643 TGGACCTCAGAGCAATATGAGCCATATATGAGCACTGAGCGGTGATTTGAGG 702
DB 1493 TGGTAAACATATGAGCAATATGAGCCATATATGAGCACTGAGCGGTGATTT 1552
QY 703 AGGATATGATTTATGAGCAATATGAGCCATATATGAGCACTGAGCGGTGATTT 762

10X

DB 1553 ATCAACCTGATTTATGAAATGCGGAGACATATGAGCATTAAGGATTTGATGATT 1612
QY 763 TCCCAAGCTACTCTTACAGATATTTGTCATATTTGAT 804
DB 1613 CTCAGATTAATTAATGATTTCAATGATGAGCATTTGAT 1654

RESULT 5
US-08-961-527-48
Sequence 48, Application US/08961527
Patent No. 642035
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 25002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-48

Query Match 4.2%; Score 37.2; DB 4; Length 25002;
Best Local Similarity 56.6%; Pred. No. 1.3;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 562 AGACTTACAGCAATATGAGCCGTTGATGATTTATATCTCTCTCTCTCTCTCTCT 621
DB 8192 AGATTTCTGGGTTTATGCTTGCAAGTCAGTATTTCTCAGAGCTTGGAGCTATTT 8251
QY 622 TTCTTATTTATCTTCAATTTTGGACCTCAGACAAATTTAGCCATTAATTTCAAC 681
DB 8252 TTCTTATTTTCTCTCAAAAAGCTCTCCAAAAGAAAGAGATTAATTCACACAC 8311
QY 682 CT 683
DB 8312 TT 8313

RESULT 6
US-08-724-394A-20
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.

APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-20

Query Match 4.2%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 49.0%; Pred. No. 3.3;
Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 434 CCCAAGCTGAACAGTAGTAGAGAGAAAAGATTGATGAGAGAAATTAAGAGCAGA 493
DB 52885 CTCAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAGGATC 52944
QY 494 AGATGATTCATAGCTCAGTAAATTTATATATTTGTATGATGATGTAACCTCTG 553
DB 52945 AGGAGTTTGATTAATTAAGTTAAAGTAGTGTTCTAGACTATTTGTCTTTATA 53004
QY 554 AATGCTGAGACTAGAGAAATGCGCTGTTGTATCATTTATATCTCTTCTTCAATT 613
DB 53005 TTAAGTGAATTTTGTAGGAGAGCATGTGTGCTCTTTTGTCTTTTGTGTT 53064
QY 614 GGCTGATTTCTTACTTATCT 635
DB 53065 GTCTGTTTTTTAATCAATCT 53086

RESULT 7
US-08-724-394A-21
Sequence 21. Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.

APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchinashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-21

Query Match 4.2%; Score 37.2; DB 2; Length 246240;
Best Local Similarity 49.0%; Pred. No. 3.3;
Matches 99; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
QY 434 CCCAAGCTGAACAGTAGTAGAGAGAAAAGATTGATGAGAGAAATTAAGAGCAGA 493
DB 52885 CTCAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAGGATC 52944
QY 494 AGATGATTCATAGCTCAGTAAATTTATATATTTGTATGATGATGTAACCTCTG 553
DB 52945 AGGAGTTTGATTAATTAAGTTAAAGTAGTGTTCTAGACTATTTGTCTTTATA 53004
QY 554 AATGCTGAGACTAGAGAAATGCGCTGTTGTATCATTTATATCTCTTCTTCAATT 613
DB 53005 TTAAGTGAATTTTGTAGGAGAGCATGTGTGCTCTTTTGTCTTTTGTGTT 53064
QY 614 GGCTGATTTCTTACTTATCT 635
DB 53065 GTCTGTTTTTTAATCAATCT 53086

RESULT 8
US-08-724-394A-22
Sequence 22. Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.

```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Charles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,395
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 94908
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3182 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-971-395-1

Query Match 4.0%; Score 35.4; DB 4; Length 3182;
Best Local Similarity 48.3%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

Qy 568 TAGCAGAAATGCGCTGTTGGTACATTATATCTCTTCCTTACGTGGCTGATTTCCTTA 627
Db 197 TAGCAATATATGACATGATTCCTAAATTGTTTGGACACCTTTTTCCTCTCTTGGTG 256

Qy 628 CTTTATCTTCAATTTTGGCAGCTCAGACGAACAATTAGCCATTAATTCACACCTGGAG 687
Db 257 TTTTCTTAACTTGAAGAAACCACTAACATGATGACGTTCAATTATTAATAAACAATATT 316

Qy 688 GGTGTGCTTTGAGGAGGATATGATTTTATGAGATGATATGGAATGCGCTTAAGA 747
Db 317 TCCAAGTTTATATACGAAACTTGTTTTTTAAATGAAGAAACAGTTGAATGATTATGA 376

Qy 748 TTTTGATGAAAAAGTTTCCCAAGCTA 772
Db 377 ATTAGTAGATCAATATCAATATATA 401

RESULT 10
US-08-413-135-1
Sequence 1, Application US/08413135
Patent No. 5689042
GENERAL INFORMATION:
APPLICANT: Amasino, Richard M
APPLICANT: Gan, Suehang
TITLE OF INVENTION: Transgenic Plants with Altered
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Charles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413.135
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.92808
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3183 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "SNG12-1 Promoter DNA"
US-08-413-135-1
```

```
Query Match 4.0%; Score 35.4; DB 1; Length 3183;
Best Local Similarity 48.3%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
```

```
QY 568 TAGCAGAAATGCGCTGTTGTACATTATATCTCTCTAGTGGCTGATTCTTA 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 TAGCTAATATAGCATGATTCATTAATTTGTTTGTACACCTTTTCTCTTTGGTG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 628 CTTATCTCTTTTGGCAGCTCAGACAAATTAAGCCATAATTCACACTGAG 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 TTTCTTAACATTAAGAAACCCATTAACAAATGATTCATTAATTAATAAACAATATT 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 688 GGTGTGTTTGGAGAGGATATGATTTTATGAGAAATGATATGGCAATGCTTAAGA 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 317 TCAGATTTTATATACGAAACTGTGTTTATATCAAAAACAGTTGATGATTATGA 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 748 TTTGATGAAAAGTTTCCCAAGCTA 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 377 ATTAGTAGATCAATCAATCAATATA 401
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
US-09-328-111-766/C
Sequence 766, Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derti, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328, 111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastsEQ for Windows Version 3.0
SEQ ID NO 766
LENGTH: 569
```

```
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(569)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-766
```

```
Query Match 4.0%; Score 35.2; DB 4; Length 569;
Best Local Similarity 52.0%; Pred. No. 0.99;
Matches 79; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
```

```
QY 382 TGCGTCAATCATCGGAATCCGAAATATTAAGTAAGTAATATGCCCCAAGC 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 TGAAGACATTAAGTTGTAATCACTGATACAGTGAAGCAATTAATTTATAGAAA 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 442 TGAACAGTAGTAGAAGAAAAAGATTTGTCAGAAATTAAGAGCAGAGATGGAT 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 TGCAATTAATCATTCATTAAGAAAGCTTTGATGAGAAATATGATAGCACTTAATAT 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 502 TCATAGCTCACTAAATTTATATATTGTA 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 ACTAGACATTAATTAATGTCATTAATGTA 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
US-08-630-822A-82
Sequence 82, Application US/08630822A
Patent No. 5840695
GENERAL INFORMATION:
```

```
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSER: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 1383 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-630-822A-82
```

```
Query Match 3.9%; Score 35; DB 2; Length 1383;
Best Local Similarity 50.3%; Pred. No. 1.6;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
```

```
QY 387 TCATCTATGGAATACAGAAATTTTAAAGTAGAATAAATATGTCCTCCAAAGCTGAGAC 446
```


DB 690 TCTAGATACGAAACGAAAGAAATATTAATCTCTGTCGTGTCAGAGATATACGAA 749
QY 447 AGTAGAGAAAGAAAGATGATGTAAGAAATTAAGAGCAGAAATGATTCAT 506
DB 750 AATTAATGATGAGAAACAAAGTCGTCAAAAGACAGAGTGCGATTAATAAGAAAGTGA 809
QY 507 AGCTCACTAAATTTATATATTTGATGATGATGTAAGCTCTGTAATG 557
DB 810 AGAACATATAAAGTGAAGATATCAAACTGATTTGCAATCTTCTCAAG 860

RESULT 13

US-09-005-069-82
Sequence 82, Application US/09005069
Patent No. 5932470
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALKENFELS, LYNDIA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005.069
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 82:
SEQUENCE CHARACTERISTICS:
LENGTH: 1383 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-005-069-82

Query Match 3.9%; Score 35; DB 2; Length 1383;
Best Local Similarity 50.3%; Pred. No. 1.6;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 387 TCAATCTATCGGAATCCAGATTTTAAAGTGAACCTAAATATGTCCTCCAAAGCTGAAC 446
DB 690 TCTAGATACGAAACGAAAGAAATTAATCTCTGTCGTGTCAGAGATATACGAA 749
QY 447 AGTAGAGAAAGAAAGATGATGTAAGAAATTAAGAGCAGAAATGATTCAT 506
DB 750 AATTAATGATGAGAAACAAAGTCGTCAAAAGACAGAGTGCGATTAATAAGAAAGTGA 809
QY 507 AGCTCACTAAATTTATATATTTGATGATGATGTAAGCTCTGTAATG 557
DB 810 AGAACATATAAAGTGAAGATATCAAACTGATTTGCAATCTTCTCAAG 860

RESULT 14

US-09-171-156A-28
Sequence 28, Application US/09171156A
Patent No. 616846
GENERAL INFORMATION:
APPLICANT: Hunter, Shirley Wu
APPLICANT: Sim, Gek-Kee
APPLICANT: Weber, Eric R.
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SHERIDAN ROSS P.C.
STREET: 1560 BROADWAY, SUITE 1200
CITY: DENVER
STATE: CO
COUNTRY: U.S.A.
ZIP: 80202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/171.156A
FILING DATE: 04-Mar-1999
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C4-PUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1383 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-171-156A-28

Query Match 3.9%; Score 35; DB 4; Length 1383;
Best Local Similarity 50.3%; Pred. No. 1.6;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 387 TCAATCTATCGGAATCCAGATTTTAAAGTGAACCTAAATATGTCCTCCAAAGCTGAAC 446
DB 690 TCTAGATACGAAACGAAAGAAATTAATCTCTGTCGTGTCAGAGATATACGAA 749
QY 447 AGTAGAGAAAGAAAGATGATGTAAGAAATTAAGAGCAGAAATGATTCAT 506
DB 750 AATTAATGATGAGAAACAAAGTCGTCAAAAGACAGAGTGCGATTAATAAGAAAGTGA 809
QY 507 AGCTCACTAAATTTATATATTTGATGATGATGTAAGCTCTGTAATG 557
DB 810 AGAACATATAAAGTGAAGATATCAAACTGATTTGCAATCTTCTCAAG 860

RESULT 15

US-08-630-822A-67
Sequence 67, Application US/08630822A
Patent No. 5840695
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS

NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-Apr-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 24..1409
US-08-630-822A-67

Query Match 3.9%; Score 35; DB 2; Length 1704;
Best Local Similarity 50.3%; Pred. No. 1.8;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 387 TCAATCTATCGGAATCCGATATTAAAGTGAATAAATATGTCCTCCCAAGCTGAAC 446
DB 713 TCTAGATACGAAACGAAAGAAATTTATTAATCTCTGCTGTCAGATATACGAA 772
QY 447 ACTAGTAGGAGAGAAAAGAGATTGATGTGAAGAAATAAAGAGGAGAGATGATTCAAT 506
DB 773 AATAATAGATGAGAAACAAAGTCGTCAAAAGACACAGAGTGCATTAATAAGAAAGTGA 832
QY 507 AGCTCACTAAATTTTATATTTGATGATGATGATGATGATGATGATGATGATGATG 557
DB 833 AGAAGCATTAAGAGTGAAGATATCAAACTGATTTTGAATCTTCTCAAG 883

Search completed: April 15, 2003, 22:46:45
Job time : 346.483 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 15, 2003, 16:52:31 ; Search time 1839.69 seconds
(without alignments)
7808.593 Million cell updates/sec

Title: US-09-647-019-3

Perfect score: 887

Sequence: 1 ggtctcctcaatccggagagag.....tgactaaatcgaatcga 887

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues 32308132

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_gatda:*
2: em_gatm:*
3: em_gatn:*
4: em_gatm:*
5: em_gatm:*
6: em_gatm:*
7: em_gatm:*
8: em_gatm:*
9: gb_gat1:*
10: gb_gat2:*
11: gb_gat3:*
12: gb_gat4:*
13: gb_gat5:*
14: gb_gat6:*
15: em_gatm:*
16: em_gatm:*
17: gb_gat7:*
18: em_gatm:*
19: em_gatm:*
20: em_gatm:*
21: em_gatm:*
22: em_gatm:*
23: em_gatm:*
24: em_gatm:*
25: em_gatm:*
26: em_gatm:*
27: em_gatm:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710.6	80.1	719	14	BM697544 UI-E-DXO-
2	669.4	75.5	683	12	BE856875 7670603.x
3	629.2	70.9	756	12	BF791178 602251278
4	598.2	67.4	909	12	BF790243 602249777
5	577	65.1	855	12	BF672902 602152759
6	575.8	64.9	793	12	BF671987 602152408

C	7	560	63.1	663	12	BF111459	BF111459 7134h12.x
C	8	554.2	62.5	906	12	BF672126	BF672126 602152580
C	9	538	60.7	879	12	BF575112	BF575112 602134792
C	10	531.8	60.0	587	9	AA211443	AA211443 zms5b01.8
C	11	518	58.4	541	14	BM717052	BM717052 UI-E-EJO-
C	12	518	58.4	544	14	BM674432	BM674432 UI-E-EJO-
C	13	495	55.8	495	9	AI090520	AI090520 q470d01.x
C	14	493	55.6	494	10	AW170011	AW170011 xj35g08.x
C	15	469.6	52.9	501	9	AA211521	AA211521 zms5b01.r
C	16	457.8	51.6	934	11	AK010172	AK010172 Mus muscu
C	17	456.2	51.4	871	12	BF693124	BF693124 602080151
C	18	452	51.0	452	9	AI190689	AI190689 qd50d01.x
C	19	450.2	50.8	614	12	BF223758	BF223758 1M0001860
C	20	449	50.6	455	9	AI355676	AI355676 q556e08.x
C	21	443.8	50.0	934	11	AK003105	AK003105 Mus muscu
C	22	438	49.4	439	9	AI333485	AI333485 qd88f03.x
C	23	428.8	48.3	890	9	AV088480	AV088480 AV088480
C	24	425.6	48.0	582	12	BG224246	BG224246 1M0030E01
C	25	397.4	44.8	544	10	BE014221	BE014221 125832 MA
C	26	396.2	44.7	826	12	BF126275	BF126275 601650456
C	27	391.8	44.2	528	12	BF073866	BF073866 220906 MA
C	28	389.2	43.9	517	12	BG224245	BG224245 1M0028F05
C	29	383	43.2	521	12	BF261282	BF261282 1M0004B05
C	30	383	43.2	867	12	BF790636	BF790636 602250354
C	31	363.2	40.9	368	14	F36800	F36800 HSPD34697 H
C	32	360	40.6	360	9	AI355905	AI355905 3055-10R-
C	33	356.8	40.2	394	9	AA386647	AA386647 M104 Retn
C	34	342.6	38.6	499	12	BG224307	BG224307 1M0033A12
C	35	330	37.2	330	9	AI355903	AI355903 BH11-F1-C
C	36	330	37.2	330	9	AI355904	AI355904 BH11-F1-C
C	37	319	36.0	466	12	BG358790	BG358790 BOWMS1-01
C	38	309.8	34.9	313	14	F26916	F26916 HSPD14595 H
C	39	308.2	34.7	480	9	AA214155	AA214155 zms5b01.0.r
C	40	307	34.6	320	14	W07478	W07478 za96c10.r1
C	41	305.2	34.4	412	12	BG223652	BG223652 1M00010F0
C	42	302.8	34.1	426	10	BE013988	BE013988 125544 MA
C	43	300.8	33.9	321	9	AI824667	AI824667 wc48c01.x
C	44	299.2	33.7	613	14	B0554133	B0554133 H4026C08-
C	45	296.4	33.4	486	9	AI035961	AI035961 ub50b12.r

ALIGNMENTS

RESULT 1
LOCUS BM697544 719 bp mRNA linear EST 28-FEB-2002
DEFINITION UI-E-DXO-agn-1-12-0-UI.r1 UI-E-DXO Homo sapiens CDNA clone
ACCESSION BM697544
VERSION BM697544.1 GI:19010802
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 719)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL (Genome Res. 6 (9), 791-806 (1996))
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.
Location/Qualifiers
source

1. 719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-DX0-agn-1-12-0-UI"
/clone_lib="UI-E-DX0"
/issue_type="Fetal eyes"
/dev_stage="Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pUT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-DX0 is a cDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pUT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGATCAAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 226 a 142 c 163 g 186 t 2 others

ORIGIN

Query Match 80.1%; Score 710.6; DB 14; Length 719;

Best Local Similarity 99.2%; Pred. No. 3.4e-150;
Matches 713; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

46 AAACGATCGAATGAGATCGACGCTCAGAGACACCGGCGCCCTTCCACTTCCAG 105
1 AAACGATCGAATGAGATCGACGCTCAGAGACACCGGCGCCCTTCCACTTCCAG 60
106 GAGCTTGTATCTTTCGATCTGGCTGCTGGGACTTCCCTTAGGCAATTAATACAT 165
61 GAGCTTGTATCTTTCGATCTGGCTGCTGGGACTTCCCTTAGGCAATTAATACAT 120
166 AAACGATCGAATGAGATCGACGCTCAGAGACACCGGCGCCCTTCCACTTCCAG 225
121 AAACGATCGAATGAGATCGACGCTCAGAGACACCGGCGCCCTTCCACTTCCAG 180
226 CCAGGCAATATCAATATTCATATTCGAGAGAGCTTCCGCGCAGAGAGAGTCAACCCCGAG 285
181 CCAGGCAATATCAATATTCATATTCGAGAGAGCTTCCGCGCAGAGAGAGTCAACCCCGAG 240
286 AAAGAAAGATGATCTCTGTAAGTGAAGAGAGAGGCTTCTCCACCTTGATGAGAGAA 345
241 AAAGAAAGATGATCTCTGTAAGTGAAGAGAGAGGCTTCTCCACCTTGATGAGAGAA 300
346 GAAGCCAAATTCAGAGAGAGAGAACTTCAGAGAGCTGCAATCTATGGAATCA 405
301 GAAGCCAAATTCAGAGAGAGAGAACTTCAGAGAGCTGCAATCTATGGAATCA 360
406 GAATATTTAAAGTGAATTAATATGTCCTCCAAAGCTGAACATGTAAGAAAG 465
361 GAATATTTAAAGTGAATTAATATGTCCTCCAAAGCTGAACATGTAAGAAAG 420
466 GATTGATGTAAGAAATTAAGAGAGAGAGTGAATTAATAGCTCACTAAATTTATA 525
421 GATTGATGTAAGAAATTAAGAGAGAGTGAATTAATAGCTCACTAAATTTATA 480
526 TATTGATGTAAGTGAAGTCTCTGAATGCTGAGACTCTAGCAGAAATGGCTGT 585
481 TATTGATGTAAGTGAAGTCTCTGAATGCTGAGACTCTAGCAGAAATGGCTGT 540
586 TGTACATTTATATCTCTCTCTAGTGGCTGTATTTCTACTTTATCTCATTTTGG 645
541 TGTACATTTATATCTCTCTCTAGTGGCTGTATTTCTACTTTATCTCATTTTGG 600

Qy 646 CACTCAGACAGCAATTAATGACCCCAATTAATTCAGACCTGAGGGTGTGTTGAGAGG 705
Db 601 CACTCAGACAGCAATTAATGACCCCAATTAATTCAGACCTGAGGGTGTGTTGAGAGG 660
Qy 706 GATATGATTTATGAGAGATGATATGAGAGAGTGTGCTTACGATTTTGAAGAAAGTTTC 764
Db 661 GATATGATTTATGAGAGATGATATGAGAGAGTGTGCTTACGATTTTGAAGAAAGTTTC 719

RESULT 2
BE856875/c
LOCUS
DEFINITION
7770C03.x1 Soares_NSF_F8_9W_OT_PA.P.S1 Homo sapiens cDNA clone IMAGE:3300004.3 similar to TR:090HP9 Q90HP9 SMALL MUSCULAR PROTEIN. ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
BE856875
BE856875.1 GI:10370341
EST.

SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 683)
NCI-CCAG <http://www.ncbi.nlm.nih.gov/ncicagp>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Straubeberg, Ph.D.
Email: cgaaps-remail.nih.gov
This clone is available royalty-free through LBNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from Gibco
High quality sequence stop: 444.

FEATURES

source

1. 683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3300004"
/clone_lib="Soares_NSF_F8_9W_OT_PA.P.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pUT73-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Benito Soares and M. Fatima Bonaldo."

BASE COUNT 198 a 142 c 123 g 220 t

ORIGIN

Query Match 75.5%; Score 669.4; DB 12; Length 683;

Best Local Similarity 99.7%; Pred. No. 6.7e-141;
Matches 681; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 200 CAGCAGTTTCCATGTTTAAAGCATTCAGGCAATATTCATATTCATATGAGAGCCTTT 259
Db 683 CAGCAGTTTCCATGTTTAAAGCATTCAGGCAATATTCATATTCATATGAGAGCCTTT 625
Qy 260 CGGCGAGAGAGGTCAACCCCGAAGAAAGATGTACTCTGTAAGTGAAGAGAGGT 319
Db 624 CGGCGAGAGAGGTCAACCCCGAAGAAAGATGTACTCTGTAAGTGAAGAGAGGT 565
Qy 320 GTTCTCCACCTCGATGAGAGAGAGAGCAATTCAGAGAGCAAGAAATTCAGAGA 379

BF790243 909 bp mRNA linear EST 12-JAN-2001
 LOCUS 602249777P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
 mRNA sequence.
 ACCESSION BF790243
 VERSION BF790243.1 GI:12095188
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 909)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1187 row: 1 column: 19
 High quality sequence stop: 626.
 Location/Qualifiers
 1..909
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4328154"
 /clone_1ib="NIH MGC 81"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site: 1: SfiI (ggcgcgcgcgcgc); Site 2: SfiI
 (ggcgcgcgcgcgc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCCGCGCGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 279 a 200 c 212 g 218 t
 ORIGIN
 Query Match 67.4%; Score 598.2; DB 12; Length 909;
 Best Local Similarity 98.8%; Pred. No. 8e-125;
 Matches 655; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
 QY 1 GGTTCGAAATCCGGAGAGGACACAGAGCTTTTCAGGCATGAAAGCATCGGAATTG 60
 DB 1 GGTTCGAAATCCGGAGAGGACACAGAGCTTTTCAGGCATGAAAGCATCGGAATTG 60
 QY 61 AGATCGAGCTCAGAGGACACCGGGCGCCCTTCACCTTCAGAGAGCTTTGATTCTT 120
 DB 61 AGATCGAGCTCAGAGGACACCGGGCGCCCTTCACCTTCAGAGAGCTTTGATTCTT 119
 QY 121 GCATCTGCGCTCGGAGACTTCCCTTAGGAGTAACAACATATACAAAGAGGATAGA 180
 DB 120 GCATCTGCGCTCGGAGACTTCCCTTAGGAGTAACAACATATACAAAGAGGATAGA 178
 QY 181 CTGATGAATATGTCGAAACGCGCAATTTCCATGTTTAGCCATCCAGGCAATATCAA 240
 DB 179 CTGATGAATATGTCGAAACGCGCAATTTCCATGTTTAGCCATCCAGGCAATATCAA 238
 QY 241 TATTCAGATGGAGCTTTGCGCCAGAGAGAGTCAACCCCGAGAGAAAAGATATAC 300
 DB 239 TATTCAGATGGAGCTTTGCGCCAGAGAGAGTCAACCCCGAGAGAAAAGATATAC 298
 QY 301 TCTGAAAGTGAGAGAGGGTGTCTCCCACTCGATGAGAGAGCAATTCAGG 360
 DB 301 TCTGAAAGTGAGAGAGGGTGTCTCCCACTCGATGAGAGAGCAATTCAGG 360

DB 299 TCTGAAAGTGAGAGAGGGTGTCTCCCACTCGATGAGAGAGCAATTCAGG 358
 QY 361 AGCGAAGAACTTCAGGAGCTCGAGTCAATCTATCGGAATCCAGAAATATTAAGTGA 420
 DB 359 AGCGAAGAACTTCAGGAGCTCGAGTCAATCTATCGGAATCCAGAAATATTAAGTGA 418
 QY 421 ACTAAATATGTCTCCCAAGCTGAACAGTACTAGAGAGAAAAGATGTAAGTGA 480
 DB 419 ACTAAATATGTCTCCCAAGCTGAACAGTACTAGAGAGAAAAGATGTAAGTGA 478
 QY 481 ATAAAGAGGAGAGATGATTAATAGCTACATAAATTTATATTTGTATAGAT 540
 DB 479 ATAAAGAGGAGAGATGATTAATAGCTACATAAATTTATATTTGTATAGAT 538
 QY 541 TGTAACTCTGTAATGCTGAGACTCTAGACAGAAATGGCCTGTTGTACA-TTTATATC 599
 DB 539 TGTAACTCTGTAATGCTGAGACTCTAGACAGAAATGGCCTGTTGTACATTTATATC 597
 QY 600 TCTTCCTTTAGTGGCTGTAATTTCTTACTTTATCTTCATTTTGGACCTCAGACACA 659
 DB 598 TCTTCCTTTAG-TGGCTGTAATTTCTTACTTTATCTTCATTTTGGGACCTCAGACACA 656
 QY 660 AAT 662
 DB 657 ATT 659
 RESULT 5
 BF672902 855 bp mRNA linear EST 21-DEC-2000
 LOCUS 602152759P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293878 5',
 mRNA sequence.
 ACCESSION BF672902
 VERSION BF672902.1 GI:11946797
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 855)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM1142 row: e column: 15
 High quality sequence stop: 687.
 Location/Qualifiers
 1..855
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4293878"
 /clone_1ib="NIH MGC 81"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site: 1: SfiI (ggcgcgcgcgcgc); Site 2: SfiI
 (ggcgcgcgcgcgc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCCGCGCGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 270 a 159 c 210 g 216 t

QY 579 GCGTGTGACATTATATATCTCTCTAGTTGGCTGATTTCTTACTTATCTCA 638
 DB 541 GCGTGTGACATTATATATCTCTCTAGTTGGCTGATTTCTTACTTATCTTC 600
 QY 639 TTTTGGACCTCAGAGAAATAGCCCATTAATTCAGACCTGAGGGGTG 693
 DB 601 ATTITGGCA-CTCAGAGAAATAGCCCATTAATTCAGACCTGAGGGGTG 654

RESULT 7
 BF11459/c 683 bp mRNA linear EST 20-OCT-2000
 LOCUS 7134h12.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:3523462.3 Similar to TR:Q9UH9 Q9UH9 SMALL MUSCULAR
 PROTEIN.; mRNA sequence.
 ACCESSION BF11459 GI:10941149
 VERSION BF11459
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 683)
 NCBI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 470.
 Location/Qualifiers
 1. 683
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3523462"
 /clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pUT73D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and 88 circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HF-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326663 Soares NBHOT
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 199 a 127 c 140 g 217 t
 ORIGIN

Query Match 63.1%; Score 560; DB 12; Length 683;
 Best Local Similarity 94.2%; Pred. No. 3.4e-116;
 Matches 603; Conservative 0; Mismatches 35; Indels 2; Gaps 2;

QY 245 CCAATGGAGCTTTGCGCCAGAGAGCTCAACCCCAAGAGAAAGATGACTCT 304
 DB 683 CCAATGGAGCTTTGCGCCAGAGAGCTCAACCCCAAGAGAAATATATGACTCT 624
 QY 305 GAATGAGAGAGGTGTTC-TCACCTCGAGAGAGAGAGAGAGCCATTCAGAGAGC 363
 DB 623 ATATGTGAGAGAGGTGTTCCTCCACATCGAGAGAGAGAGAGAGCCATTCAGAGAGC 564
 QY 364 GAAAGAACTTCAGAGAGCTGAGTCAATCTATGGAATATCAAGAAATTTAAAGTGA 423

DB 563 GAAAGAACTTCAGAGAGCTGAGTCAATCTATGGAATATCAAGAAATTTAAAGTGA 504
 QY 424 AAAATATGTCCTCCCAAGCTGAACAGTAGAGAGAGAGAGAGAGAGAGAGAGATA 483
 DB 503 AAAATATGTCCTCCCAAGCTGAACAGTAGAGAGAGAGAGAGAGAGAGAGAGATA 444
 QY 484 AAGAGGAGAGAGATGATTCATATGCTCATAAATTTATATTTGTATGATGATGT 543
 DB 443 AAGAGGAGAGAGATGATTCATATGCTCATAAATTTATATTTGTATGATGATGT 384
 QY 544 GAACCTCCTGAATGCTGAGAGCTTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 603
 DB 383 GAACCTCCTGAATGCTGAGAGCTTAGCAGAGAGAGAGAGAGAGAGAGAGAGATA 324
 QY 604 CTTTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGGACCTCAGAGAGAGATA 663
 DB 323 CTTTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGGACCTCAGAGAGATA 264
 QY 664 AGCCATTAATTCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 723
 DB 263 AGCCATTAATTCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 204
 QY 724 ATGATATGATA 782
 DB 203 ATGATATGATA 144
 QY 783 TATTTGGTCAATATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 842
 DB 143 TATTTGGTCAATATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATA 84
 QY 843 CTTTGTATGAGTCAATTAATTTTGTACCTTAATGTAATA 882
 DB 83 CTTTGTATGAGTCAATTAATTTTGTACCTTAATGTAATA 44

RESULT 8
 BF672126 906 bp mRNA linear EST 21-DEC-2000
 LOCUS 602152580F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293763.5,
 DEFINITION mRNA sequence.
 ACCESSION BF672126
 VERSION BF672126 GI:11946021
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 906)
 NIH-MGC http://mgc.ncbi.nlm.nih.gov/
 TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LNCM141 row: P column: 20
 High quality sequence stop: 621.
 Location/Qualifiers
 1. 906
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4293763"
 /clone_id="NIH_MGC_81"
 /lab_host="DH10B (71 phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site 1: SfiI (ggccgctggcc); Site 2: SfiI
 (ggccattatggcc); 5' and 3' adaptors were used in cloning

Query Match	60.0%	Score	531.8	DB	9	Length	587
Best Local Similarity	97.1%	Pred. No.	7.9e-110				
Matches - 572; Conservative	0	Mismatches	14	Indels	3	Gaps	3

Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).

Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES
Source
1. .541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UI-E-EJ0-ahk-c-03-0-UI"
/issue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: Ecor I; Site 2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGATCAAGA
; lens, CCAATAGCA; eye anterior segment, AATGCCCAT;
optic nerve, CCAATAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT
188 a 82 c 100 g 171 t

Query Match 58.4%; Score 518; DB 14; Length 541;
Best Local Similarity 100.0%; Pred. No. 1e-106;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
ORIGIN
370 ACTTCGAGACCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 429
1 ACTTCGAGACCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 60
430 TGTCCCAAGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 489
61 TGTCCCAAGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 120
490 CAGAAATGATTCATAGTCACTAAATTTATATTTGATGATTTGAACCT 549
121 CAGAAATGATTCATAGTCACTAAATTTATATTTGATGATTTGAACCT 180
550 CCGAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 609
181 CCGAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 240
610 AGTTGGCTGATTTCTACTTATCTCATTTTGGCACTTCAGAAATTAAGCCCA 669
241 AGTTGGCTGATTTCTACTTATCTCATTTTGGCACTTCAGAAATTAAGCCCA 300
670 TAAATTCACACCTGAGGCTGTTTGGAGAGATATGATTTATGAGAAATGATA 729
301 TAAATTCACACCTGAGGCTGTTTGGAGAGATATGATTTATGAGAAATGATA 360
730 TGGCAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 789
361 TGGCAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 420
790 GTCAATTTTGGAAATGGCTTTAGTTCTTCACTTTTAAATTAATGATCACTTAATTTTGA 849
421 GTCAATTTTGGAAATGGCTTTAGTTCTTCACTTTTAAATTAATGATCACTTAATTTTGA 480
850 TGAGTTCAATTAATATTTGATCACTTAATGATCACTTAATTTTGA 887
481 TGAGTTCAATTAATATTTGATCACTTAATGATCACTTAATTTTGA 518

RESULT 12
BM674432/c
LOCUS
DEFINITION
UI-E-EJ0-ahk-c-03-0-UI.61 UI-E-EJ0 Homo sapiens cDNA clone
BM674432
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 544)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.research.com).
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
Source
Location/Qualifiers

1. .544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="UI-E-EJ0-ahk-c-03-0-UI"
/issue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: Ecor I; Site 2: Not I;
UI-E-EJ0 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an Ecor I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGATCAAGA
; lens, CCAATAGCA; eye anterior segment, AATGCCCAT;
optic nerve, CCAATAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI).
TAG LIB=UI-E-EJ0
TAG_LIB=UI-E-EJ0
TAG_LIB=UI-E-EJ0
TAG_SEQ=GTCC

BASE COUNT
171 a 103 c 82 g 188 t
ORIGIN
370 ACTTCGAGACCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 429
1 ACTTCGAGACCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 60
430 TGTCCCAAGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 489
61 TGTCCCAAGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 120
490 CAGAAATGATTCATAGTCACTAAATTTATATTTGATGATTTGAACCT 549
121 CAGAAATGATTCATAGTCACTAAATTTATATTTGATGATTTGAACCT 180
550 CCGAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 609
181 CCGAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 240
610 AGTTGGCTGATTTCTACTTATCTCATTTTGGCACTTCAGAAATTAAGCCCA 669
241 AGTTGGCTGATTTCTACTTATCTCATTTTGGCACTTCAGAAATTAAGCCCA 300
670 TAAATTCACACCTGAGGCTGTTTGGAGAGATATGATTTATGAGAAATGATA 729
301 TAAATTCACACCTGAGGCTGTTTGGAGAGATATGATTTATGAGAAATGATA 360
730 TGGCAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 789
361 TGGCAATGCTGAGTCAATCTATGGAATCCAGAAATTTAAAGTGAATTAATA 420
790 GTCAATTTTGGAAATGGCTTTAGTTCTTCACTTTTAAATTAATGATCACTTAATTTTGA 849
421 GTCAATTTTGGAAATGGCTTTAGTTCTTCACTTTTAAATTAATGATCACTTAATTTTGA 480
850 TGAGTTCAATTAATATTTGATCACTTAATGATCACTTAATTTTGA 887
481 TGAGTTCAATTAATATTTGATCACTTAATGATCACTTAATTTTGA 518

[illegible]

Db	136	GGAGGAGTGTGGTTTGAAGAGGAGATATGATTTTANGAGAGATGATATGGCAATGTGGCTTA	136
Qy	744	ACGATTTTGATGAAAAAGTTTCCCAAGCTACTTCTCAACAGTATTTTGGTCATATTTGGAA	803
Db	135	ACGATTTTGATGAAAAAGTTTCCCAAGCTACTTCTCAACAGTATTTTGGTCATATTTGGAA	76
Qy	804	TGCGTTTAGTCTTCACCTTTTAATATATGCTACATAACTTTGTATGAGTTCAAAATPAA	863
Db	75	TGCGTTTAGTCTTCACCTTTTAATATATGCTACATAACTTTGTATGAGTTCAAAATPAA	16
Qy	864	TATTTGACTAAATGT 878	
Db	15	TATTTGACTAAATGT 1	

RESULT 14
 AM170011/c
 LOCUS
 DEFINITION
 xj35g98.x1 Soares NFL T GBC S1 Homo sapiens CDNA clone
 IMAGE:2659262 3', mRNA sequence.
 ACCESSION
 AM170011
 VERSION
 AM170011.1 GI:6401536
 KEYWORDS
 EST.
 SOURCE
 human.

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 494)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.
	Email: cgaps-rt@mail.nih.gov
	This clone is available royalty-free through LANT ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
	Seq primer: -40UP from Gibco.
	High quality sequence stop: 449.

OY 508 GCTCACTAAATTTTATATTTGTATGATGATGGAACCTCT-GAATGCTGAGACT 566
DB 420 GCTCACTAAATTTTATATTTGTATGATGATGGAACCTCTGGAATGCTGAGACT 479

OY 567 CTAGCAGAAATGGCCTGTTT 586
DB 480 CTAGCAGAAATGGCCTGTT 499

Search completed: April 15, 2003, 22:02:27
Job time : 1845.69 secs


```

FT      /*tag= c
FT      /product= "Chisel (Csl) protein"
FT      /note= "Expressed predominantly in heart muscles"
FT      230..316
FT      exon
FT      /tag= d
FT      /label= Exon_3
FT      /note= "Corresponds to residues 5220-5134 of
FT      human cosmid clone U28D4"
FT      317..465
FT      exon
FT      /tag= e
FT      /label= Exon_4
FT      /note= "Corresponds to residues 35384-35236 of
FT      human cosmid clone U112E8"
FT      466..887
FT      exon
FT      /tag= f
FT      /label= Exon_5
FT      /note= "Corresponds to residues 4101-3680 of
FT      human cosmid clone U112E8"
FT
FT      W09950410-A1.
FT
FT      07-OCT-1999.
FT
FT      26-MAR-1999; 99WO-AU00220.
FT
FT      27-MAR-1998; 98AU-0002634.
FT
FT      (CHAN-) CHANG CARDIAC RES INST VICTOR.
FT      (GEO) GEN HOSPITAL CORP.
FT      (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
FT
FT      Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
FT      WPI; 1999-610852/52.
FT      P-PSDB; AAY28651.
FT
FT      Isolated nucleic acids encoding chisel, used to develop products for
FT      treating cardiomyopathy, cardiac hypertrophy, heart failure and
FT      muscular myopathies -
FT
FT      Claim 8; Page 149-150; 157p; English.
FT
FT      The present sequence is the cDNA encoding the human chisel gene (csl)
FT      that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.
FT      It is a member of the EF-Hand protein super family and is involved in
FT      signalling pathways. It is predominantly expressed in heart and skeletal
FT      muscles and is activated after the differentiation of cells. Csl
FT      functions in regulation aspects of differentiation or adaptive processes
FT      that maintain muscle homeostasis. This sequence can be used in the
FT      detection, diagnosis, prophylactic and therapeutic treatment of diseases
FT      such as those involving aberrant muscle cell development and functional
FT      activity. It is also used in the treatment of muscular and myotonic
FT      dystrophies, skeletal myopathies such as Duchenne muscular dystrophy and
FT      Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,
FT      myocarditis, myofiber atrophy, etc. The Csl gene sequence can also be
FT      used in gene therapy, for the production of transgenic animals and for
FT      drug screening.
FT
FT      Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;
FT
FT      Query Match      100.0%; Score 887; DB 20; Length 887;
FT      Best Local Similarity 100.0%; Pred. No. 1.4e-238;
FT      Matches 887; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB      121 GCATCTGCTGCTGGAGCTTCCCTTAGGAGTAACAATATCAATGAAGCGAGATAGA 180
DB      181 CTGATGAAATATGTGCAAAACAGCAGTTTCCATGTTAGAGCCATCCAGCAATATCAA 240
DB      181 CTGATGAAATATGTGCAAAACAGCAGTTTCCATGTTAGAGCCATCCAGCAATATCAA 240
DB      241 TATTCGAATGGAGCTTTGGCCAGAGACAGTCAACCCCGAGAAAAGAAATGTAC 300
DB      241 TATTCGAATGGAGCTTTGGCCAGAGACAGTCAACCCCGAGAAAAGAAATGTAC 300
DB      301 TCCGAAGTGGAGAGGGGTGTTCTCCCACTCGAGTGAAGAGCAATCCAGG 360
DB      301 TCCGAAGTGGAGAGGGGTGTTCTCCCACTCGAGTGAAGAGCAATCCAGG 360
DB      361 AGCGAAGAACTTCAGAGCTCGAGTCAATCTATCCGAAATCCAGAAATTAAGTGA 420
DB      361 AGCGAAGAACTTCAGAGCTCGAGTCAATCTATCCGAAATCCAGAAATTAAGTGA 420
DB      421 ACTAAATATGTCCCAAGCTGAACAGTATAGAGAGAGAGAGAGAGAGAGAGAG 480
DB      421 ACTAAATATGTCCCAAGCTGAACAGTATAGAGAGAGAGAGAGAGAGAGAGAG 480
DB      481 ATAAAGAGGAGAGATGATGATCAATAGCTCACTAAATTTATATTTGATGATGAT 540
DB      481 ATAAAGAGGAGAGATGATGATCAATAGCTCACTAAATTTATATTTGATGATGAT 540
DB      541 TGTGAACCTCGTGAATCTGAGACTGACAGTACAGAAATGCGCTTTGATATATATCT 600
DB      541 TGTGAACCTCGTGAATCTGAGACTGACAGTACAGAAATGCGCTTTGATATATATCT 600
DB      601 CTTCCTTCTAGTGGCTGATATTTCTTATCTTATCTTATCTTATCTTATCTTATCT 660
DB      601 CTTCCTTCTAGTGGCTGATATTTCTTATCTTATCTTATCTTATCTTATCTTATCT 660
DB      661 ATTAGCCATTAATTAATCAACCTGAGAGGGTGGTTTGAAGAGGATATGATTTATGG 720
DB      661 ATTAGCCATTAATTAATCAACCTGAGAGGGTGGTTTGAAGAGGATATGATTTATGG 720
DB      721 AGAATGATATGGCAATGTGCTAAGATTTGATGAGAAAGTTTCCAGTACTTCTTAC 780
DB      721 AGAATGATATGGCAATGTGCTAAGATTTGATGAGAAAGTTTCCAGTACTTCTTAC 780
DB      781 AGATTTTGTCAATATTTGGAATGCGTTTATGTTCTTCACTTTAAATATATGTCATA 840
DB      781 AGATTTTGTCAATATTTGGAATGCGTTTATGTTCTTCACTTTAAATATATGTCATA 840
DB      841 AACTTGTATGAGTTCAATTAATATTTGACTAAATGTAAATGTA 887
DB      841 AACTTGTATGAGTTCAATTAATATTTGACTAAATGTAAATGTA 887

RESULT 2
AADD27216
ID AADD27216 standard; DNA; 886 BP.
XX
XX AAD27216;
AC
XX
XX 09-APR-2002 (first entry)
DT
XX
XX Human 66214 EST clone DNA.
DE
XX
XX Human; congestive heart failure; dilative cardiomyopathy; sudden death;
XX hypertrophic cardiomyopathy; ischemic cardiomyopathy; rhythm disorder;
XX heart muscle disease; conduction disorder; coronary heart disease;
XX systemic arterial hypertension; pulmonary hypertension; endocarditis;
XX pulmonary heart disease; valvular heart disease; pericardial disease;
XX congenital heart disease; gene therapy; syncope; transgenic animal;
XX expressed sequence tag; EST; clone 66214; de.
XX Homo sapiens.
XX
XX Key Location/Qualifiers

```


PT detecting expression levels of one or more genes in prostate cells from
PT patient that are differentially regulated compared to normal prostate
PT cells -

PS Disclosure; Page 104; 444pp; English.

CC The invention relates to a method of diagnosing (I) the onset or
CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
CC or identifying an agent that modulates the onset or progression of BPH.
CC The method is based on changes in gene expression in BPH tissue
CC from patients exhibiting different clinical states of prostate
CC hyperplasia as compared to normal prostate tissue. (I) comprises
CC detecting the expression levels of one or more genes in prostate cells
CC from the subject that are differentially regulated compared to normal
CC prostate cells. (II) comprises preparing a first gene expression profile
CC of BPH cells or BPH-like cell population, exposing the cells to the
CC agent, preparing a second gene expression profile of the agent exposed
CC cells, and comparing the first and second gene expression profiles.
CC (I) is useful for diagnosing the onset or progression of BPH. (II) is
CC useful for identifying an agent that modulates the onset or progression
CC of BPH. The methods are useful to prevent information identifying
CC the expression level in a tissue or cells. By comparing the expression
CC level of genes given in the specification in the tissue or cells to the
CC level of expression of at least one gene in the database, and displaying the
CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK4106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.

Sequence 587 BP; 177 A; 122 C; 103 G; 183 T; 2 other;

Query Match	60.0%;	Score 531.8;	DB 24;	Length 587;
Best Local Similarity	97.1%;	Pred. No. 5.1e139;		
Matches 572;	Conservative	0;	Mismatches 14;	Indels 3;
				Gaps 3;

QY	290	AAAGAATGTACTCTCGAAGTGTGAGAGGGTCTCTCCACCTCGAGTGAAGGAGAAAG	349
Db	587	AAAAGAATGTCTCCCGAAGTGTGAGAGGGTCTCTCCACCTCGAGTGAAGGAGAAAG	529
QY	350	CCAATTCCAGAGCGAGAGAACTTCAGAGCTTCAGTCAATCTATCGGAATCCAGAT	409
Db	528	CCAATTCCAGAGCGG-AGGAACCTCCGAGNCTCAGTCAATCTATCGGAATCCAGAT	470
QY	410	ATTAAAGTGA-CTAAATATGTGCCCAAGCTGAACAGTATGAGAGAAAAAGAT	468
Db	469	ATTAAAGTGAACCTAAATATGTGCCCAAGCTGAACAGTATGAGAGAAAAAGAT	410
QY	469	TGATGTGAAGAAATGAAGAGCGAGCAAGATGATTCATATGCTCACTAAATTTATAT	528
Db	409	TGATGTGAAGAAATGAAGAGCGAGCAAGATGATTCATATGCTCACTAAATTTATAT	350
QY	529	TTGTATGATGATTTGTGAACCTCTCGAATGCTGAGACTCTAGCAGAAATGCGCTGTTGT	588
Db	349	TTGTATGATGATTTGTGAACCTCTCGAATGCTGAGACTCTAGCAGAAATGCGCTGTTGT	290
QY	589	ACATTATATCTCTCCCTTCAGTGTGATGATTTCTACTTTATCTCAATTTTGGCAC	648
Db	289	ACATTATATCTCTCTTCCTTCAGTGTGATGATTTCTACTTTATCTCAATTTTGGCAC	230
QY	649	CTCAGAGACAAATTAGCCCTAAATTTCAACACCTGAGGGGTGTGTTTGAAGAGGAT	708
Db	229	CTCAGAGACAAATTAGCCCTAAATTTCAACACCTGAGGGGTGTGTTTGAAGAGGAT	170
QY	709	ATGATTTTATGAGAAATGATATGGCAATGTGCTTAACAGATTTTGATGAAGAATTTCCCA	768
Db	169	ATGATTTTATGAGAAATGATATGGCAATGTGCTTAACAGATTTTGATGAAGAATTTCCCA	110
QY	769	GCTACTTCTACAGTATTTTGGTCAATATTTGGAATCGTTTATAGTTCTTCACCTTTAA	828
Db	109	GCTACTTCTACAGTATTTTGGTCAATATTTGGAATCGTTTATAGTTCTTCACCTTTAA	50
QY	829	ATTATGTCACTAACTTTGATGATTCGAATTAATTTTGACTAATG	877

XX WPI, 2002-188264/24.
 PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 PS Claim 1, SEQ ID 2927, 44bp; English.
 CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC osteophagel, ovarian, kidney, prostate or pancreatic cancer.
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumor.
 XX
 SQ Sequence 587 BP, 177 A, 122 C, 103 G, 183 T, 2 other:
 Query Match 60.0%; Score 531.8; DB 24; Length 587;
 Best Local Similarity 97.1%; Pred. No. 5,1e-139;
 Matches 572; Conservative 0; Mismatches 14; Indels 3; Gaps 3;
 Oy 290 AAGAATCTACTCTGGAAGTGGAGAGGGTGTCTCCACCTGGATGAGAGAAAG 349
 Db 587 AAAAGATGTCCTCCGAGAGTGGAGAGGGTG-TCTCCACCTGGATGAGAGAAAG 529
 Oy 350 CCAATTCAGAGGAGAAACCTTCAGAGCTGAGTCAATCTATCGGAATCCAGAT 409
 Db 528 CCAATTCAGAGGAGC-AAGAACCTCCGAGCTGAGTCAATCTATCGGAATCCAGAT 470
 Oy 410 ATTAAAGTGA-CTAAATATATGTCCTCCAAAGCTGACAGTAGTAGAAGAAAAAGAT 468
 Db 469 ATTAAAGTGA-CTAAATATATGTCCTCCAAAGCTGACAGTAGTAGAAGAAAAAGAT 410
 Oy 469 TGATGTGAAGAAATAAAGGCAAGATGATTCATATGCTCAATAAATTTTATAT 528
 Db 409 TGATGTGAAGAAATAAAGGCAAGATGATTCATATGCTCAATAAATTTTATAT 350
 Oy 529 TTGATGATGATGGAAGCTCTGAAATGCTGAGACTCTAGCAGAAATGGCTTTGT 588
 Db 349 TTGATGATGATGGAAGCTCTGAAATGCTGAGACTCTAGCAGAAATGGCTTTGT 290
 Oy 589 ACATTATATCTCTTCTTCTAGTGGCTGATATCTTACTTATCTTCAATTTTGGAC 648
 Db 289 ACATTATATCTCTTCTTCTAGTGGCTGATATCTTACTTATCTTCAATTTTGGAC 230
 Oy 649 CTCAGAGAAATTAAGCCATTAATTCAGACCTGAGGGGTGTTTGAAGAGGAT 708
 Db 229 CTCAGAGAAATTAAGCCATTAATTCAGACCTGAGGGGTGTTTGAAGAGGAT 170
 Oy 709 ATGATTTTATGAGAGATGATGAGCAATGTGCTTAAAGATTTTGAAGAAATTTCCCA 768
 Db 169 ATGATTTTATGAGAGATGATGAGCAATGTGCTTAAAGATTTTGAAGAAATTTCCCA 110
 Oy 769 GCTACTCTCTACAGATATTTGGTCAATATTTGAAATGCGTTTAACTCTTTTAA 828
 Db 109 GCTACTCTCTACAGATATTTGGTCAATATTTGAAATGCGTTTAACTCTTTTAA 50
 Oy 829 ATTATGCTACTAAAGTGTATGATGATCAATTAATTTTGAAGTAAATG 877
 Db 49 ATTATGCTACTAAAGTGTATGATGATCAATTAATTTTGAAGTAAATG 1

RESULT 5
 ID AAX90903
 AC AAX90903;
 DT 17-JAN-2000 (first entry)
 XX
 DS cDNA encoding murine chisel (Csl) gene.
 XX
 KW Chisel gene; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; murine;
 KW X chromosome; regulation; adaptive process; muscle homeostasis;
 KW detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KW differentiation; gene therapy; transgenic animal; drug screening; ss.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 XX CDS 199..456
 XX FT /tag= a
 XX FT /product= "Chisel (Csl) protein"
 XX FT /note= "Expressed especially in heart muscles"
 XX
 XX MO950410-A1.
 XX
 XX 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WC-AU00220.
 XX
 XX 27-MAR-1998; 98AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 XX (GEO) GEN HOSPITAL CORP.
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenchal NA;
 XX WPI; 1999-610852/52.
 XX P-PSDB; AAY28650.
 XX
 XX Isolated nucleic acid encoding chisel, used to develop products for
 XX treating cardiomyopathy, cardiac hypertrophy, heart failure and
 XX muscular myopathies -
 XX
 XX Claim 3; Page 148; 157bp; English.
 CC The present sequence is the cDNA encoding the murine chisel gene (Csl)
 CC that is mapped to the mouse X chromosome. It is a member of the EF-Hand
 CC protein super family and is involved in signalling pathways. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 SQ Sequence 778 BP, 231 A, 166 C, 179 G, 202 T, 0 other:
 Query Match 43.9%; Score 389.8; DB 20; Length 778;
 Best Local Similarity 73.5%; Pred. No. 4e-99;
 Matches 562; Conservative 0; Mismatches 152; Indels 51; Gaps 3;
 Oy 5 CTCATACCGGAGAGCAGACAGACTTTTCAGCAGATGAAAAGCATCGGAATTGAT 64

XX Human; congestive heart failure; dilative cardiomyopathy; sudden death;
 KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
 KW heart muscle disease; conduction disorder; coronary heart disease;
 KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
 KW pulmonary heart disease; valvular heart disease; pericardial disease;
 KW congenital heart disease; gene therapy; syncope; transgenic animal;
 KW clone 66214; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200192567-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001MO-EP06165.
 XX
 PR 30-MAY-2000; 2000US-207400P.
 XX
 PA (MED1-) MEDIGENE AG.
 XX
 PI Bunk D, Reuner B, Beck J, Henkel T;
 XX
 DR WPI; 2002-122073/16.
 XX
 PT Identifying a subject at risk for a heart disease e.g. congestive heart
 PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 PT the polypeptide expressed by genes abnormally expressed in heart tissue
 XX
 XX Example 10; Fig 9a; 154pp; English.
 XX
 CC The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present sequence is a cDNA from 66214 clone. This
 CC sequence used in the exemplification of the invention.
 XX
 SQ Sequence 290 BP; 103 A; 50 C; 69 G; 68 T; 0 other;
 XX
 Query Match 32.7%; Score 290; DB 24; Length 290;
 Best Local Similarity 100.0%; Pred. No. 2.5e-71;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 299 ACTCTGAAGTGAAGGAGGTGTCTCCACCTCGATGAGAGAGAACCAATTCCA 358
 Db 1 ACTCTGAAGTGAAGGAGGTGTCTCCACCTCGATGAGAGAGAACCAATTCCA 60
 QY 359 GGAGCGAAGAACTTCCAGACCTGCACTCATTCGGAATTCAGATATTTAAAGT 418
 Db 61 GGAGCGAAGAACTTCCAGACCTGCACTCATTCGGAATTCAGATATTTAAAGT 120
 QY 419 GAATTAATATATGTCCCAAGCTGAACAGATGAGAGAGAGAGATGATGAG 478
 Db 121 GAATTAATATATGTCCCAAGCTGAACAGATGAGAGAGAGATGATGAG 180
 QY 479 AATATTAAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 538
 Db 181 AATATTAAGGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 240
 QY 539 ATTGTGAACCTCTGGAATGCTGAGACTCTTGCGAAGAAATGCGCTGTTGT 588

Db 241 ATTGTGAACCTCTGGAATGCTGAGACTCTAGACCTAGCAGAAATGCGCTGTTGT 290
 RESULT 8
 ABN49389
 ID ABN49389 standard; DNA; 60 BP.
 XX
 AC ABN49389;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human spliced transcript detection oligonucleotide SEQ ID NO:22137.
 XX
 KW Human; mouse; rat; splice transcript; detection; RNA transcript;
 KW splice variant; transcriptome; oligonucleotide library; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200210449-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 20-JUL-2001; 2001MO-IB01903.
 XX
 PR 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX
 PA (COMP-) COMPUGEN INC.
 XX
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX
 DR WPI; 2002-257383/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes
 XX
 XX Example 1; SEQ ID 22137; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridising selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterising the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 60 BP; 16 A; 12 C; 14 G; 18 T; 0 other;
 XX
 Query Match 6.8%; Score 60; DB 24; Length 60;
 Best Local Similarity 100.0%; Pred. No. 4.2e-07;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 533 ATATGATTTGAACTCTGGAATGCTGAGACTCTAGCAGAAATGCGCTGTTGATCAT 592

```

Db      1  ATGATGATTTGTGAACCTCCTGAAATGCTGTAGACTTACAGAGAAATGCGCTTTTGTACAT 60
RESULT 9
ID      ABNS58272
AC      ABNS58272
XX      ABNS58272;
XX      15-JUL-2002 (first entry)
XX      Mouse spliced transcript detection oligonucleotide SEQ ID NO:31020.
DE      Human; mouse; rat; splice transcript; detection; RNA transcript;
XX      splice variant; transcriptome; oligonucleotide library; ss.
OS      Mus musculus.
PN      WO200210449-A2.
XX      07-FEB-2002.
PD
XX      20-JUL-2001; 2001WO-1B01903.
XX      28-JUL-2000; 2000US-221607P.
PR      02-MAY-2001; 2001US-287724P.
XX      (COMP-) COMPUGEN INC.
XX      Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
PI      WPI, 2002-257383/30.
XX      New oligonucleotide libraries comprising oligonucleotides which
PT      selectively hybridize to mRNAs transcribed from a transcription unit of
PT      a genome, useful for detecting tissue-, pathology-, and
PT      developmental-specific genes
XX
XX      Example 1; SEQ ID 31020; 47bp; English.
XX
XX      The present invention describes oligonucleotide libraries for detecting
XX      messenger RNAs that populate a (sub-)transcriptome, where the
XX      (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX      transcription units that populate a genome. The library comprises
XX      several oligonucleotides, each capable of hybridizing selectively to a
XX      set of messenger RNAs transcribed from a given transcription unit of
XX      the genome, which encodes one or more messenger RNA splice variants.
XX      The oligonucleotide libraries are useful for detecting mRNAs from a
XX      biological sample, in expression profiling studies, in qualitatively or
XX      quantitatively characterizing the corresponding transcriptome, and in
XX      detecting RNA transcripts and splice variants of human or animal
XX      transcriptomes. The libraries may also be used as specialised mini
XX      libraries to detect transcripts of a sub-transcriptome under a
XX      particular biological or pathological state, and so allowing the
XX      detection of tissue- and pathology-specific genes such as those genes
XX      only expressed in specific tissue under a specific pathological
XX      condition; to detect developmental specific genes; and to detect RNA
XX      transcripts and splice variants of a transcriptome of a patient suffering
XX      from a particular disorder. ABN27253 (to ABNS9569) represent
XX      oligonucleotide sequences from rats, humans and mice, which are used in
XX      the exemplification of the present invention.
XX      N.B. The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in electronic format directly from WIPO
XX      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 65 BP; 24 A; 10 C; 13 G; 18 T; 0 other;
SQ
Query Match      5.3%; Score 47.4; DB 24; Length 65;
Best Local Similarity 83.1%; Pred. No. 0.0015;
Matches 54; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
364 GAAGAACTTCAGAGACCTGCAGTCAATCTTATCGAAATCCGAAATATTTAAAGTGAAGT 423
|||||

```

Db	1	GAGGAATTTCCAGACCTGTTTCACACTTGTCTGAGATCCAAATGTTAAATGTAACACT	60
Oy	424	AAAAT 428	
Db	61	GAAAT 65	
RESULT 10			
ABL33800			
ID	ABL33800	standard; DNA; 6115 BP.	
AC			
XX	ABL33800;		
XX			
XX	26-MAR-2002	(first entry)	
DT			
DE			
XX			
XX	Human immune system associated gene SEQ ID NO: 1773.		
KW	Human; immune system disease; cytosine methylation; antiasthmatic;		
KW	antiarteriosclerotic; anti-anemic; cytosolic; noctropic;		
KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KW	antirheumatic; antiarthritis; antidiabetic; antipsoriatic;		
KW	antifibromatous; cancer; eye disease; arteriosclerosis; anaemia;		
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;		
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;		
KX	gene; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WC000200928-A2.		
XX			
XX	03-JAN-2002.		
XX			
PF	02-JUL-2001; 2001WO-EP07537.		
XX			
PR	30-JUN-2000; 2000DE-1032529.		
PR	01-SEP-2000; 2000DE-1043826.		
XX			
PA	(EPIG-) EPIGENOMICS AG.		
XX			
PI	Olek A, Piepenbrock C, Berlin K;		
XX			
PI	WPI; 2002-130909/17.		
DR			
XX			
PT	Nucleic acid comprising fragment of chemically modified gene, useful		
PT	for diagnosis and treatment of diseases associated with abnormal		
PT	cytosine methylation		
XX			
XX			
PS	Claim 1; SEQ ID NO 1773; 32pp + Sequence Listing; German.		
XX			
CC	The present invention provides a number of human immune system associated		
CC	genes which are modified by the methylation of cytosines. The sequences		
CC	can be used in the diagnosis and treatment of immune system disorders,		
CC	including eye diseases such as retinopathy, neovascular glaucoma and		
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid		
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,		
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel		
CC	diseases. The present sequence is a gene of the invention.		
XX			
XX			
XX	Sequence 6115 BP; 1623 A; 47 C; 1519 G; 2926 T; 0 other;		
Query Match	4.8%; Score 43; DB 24; Length 6115;		
Best Local Similarity	59.5%; Pred. No. 0.19;		
Matches 110; Conservative	0; Mismatches 70; Indels 5; Gaps 2;		
Oy	703	AGGATATGATTTTATGAGAAATGATATGGCAATGTCCTAAGCATTTTGATGAAAATT	762
Db	5256	AGAGTTAGTGAATGATGAGAGAGAAATGGAAATGCTGATTTTATTAATTTAA--TT	5313
Oy	763	TCCCAAGCTACTCTCTCAAGATTTTGTCATATTTTGAATGGCTTTAGTCTTACAC	822
Db	5314	TTAGCTTTTATTTTATTTATTTATTTTGGTAAATTTTGGTATGATTTTGGGCTTTTATT	5373
Oy	823	TTTTAAATATGATGACAAATCTTGTATGAGTTCAAATATATTTGATCACTAAATGTAAAA	882

DB 5374 TTTTGTGTTTATTT---TAAAGTTTGCTATTATTAAGTATTTTGAAGTATAGAA 5430
 Oy 883 TGTGA 887
 DB 5431 AAGA 5435

RESULT 11

AA567967
 ID AA567967 standard, cDNA, 1134 BP.

AC AA567967;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #3771.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN MO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABC03780.

PT New isolated polynucleotide and encoded polypeptide, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

PS Claim 1; SEQ ID NO 3771; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AA564197-AA594564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1134 BP; 325 A; 365 C; 264 G; 180 T; 0 other;

Query Match 4.6%; Score 41.2; DB 23; Length 1134;

Best Local Similarity 51.1%; Pred. No. 0.29;
 Matches 97; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Oy 289 AAAGAAATGTACTCTGAGAGAGAGGTTCTCCCACTCGATGAGAGAA 348
 DB 246 AAGGAAGGAGTTGCGAAGTGAGAGAGAGAGGATCCACGTGTGAGCAGGCCGAA 305
 Oy 349 GCCAATTCAGAGAGCAAGAACTTCAGGACTGCAATCTATCGAAATCCGAA 408
 DB 306 CAGGAAGCCAGGGGCGAAGACCAAGCAGCGCGCAGCCAAAGGGGGGAGAAACAGCG 365
 Oy 409 TATTAAAGTGAACTAAATATGTCTCCCAAGCTGAACACTAGTAGAGAAAGAAATGAT 468
 DB 366 GAAATCCGCCGAGCAAGAAAGTAGACCAAGATGTGAAGAACTCAAAAGACAGACAGAG 425
 Oy 469 TGATGTGAG 478
 DB 426 AGGGGCCGAG 435

RESULT 12

AAH24065
 ID AAH24065 standard; DNA; 4590 BP.

AC AAH24065;

DT 29-AUG-2001 (first entry)

DE Yeast AOD9604-associated DNA sequence, SEQ ID NO:1.

KW Human growth hormone analogue peptide; hGH; AOD9604; lipid metabolism;
 KW modulation; lipolysis stimulation; hormone-sensitive lipase stimulation;
 KW lipogenesis inhibition; acetyl CoA carboxylase inhibition; obesity;
 KW functional food; transgenic yeast; fat/lean ratio; food use; ds.

OS Saccharomyces cerevisiae.

XX Key Location/Qualifiers

FT misc_feature 10

FT /tag= a
 FT /note= "Represented as * in the specification"

FT /tag= b
 FT /note= "Represented as * in the specification"

FT /tag= c
 FT /note= "Represented as * in the specification"

FT /tag= d
 FT /note= "Represented as * in the specification"

FT /tag= e
 FT /note= "Represented as * in the specification"

FT /tag= f
 FT /note= "Represented as * in the specification"

FT /tag= g
 FT /note= "Represented as * in the specification"

FT /tag= h
 FT /note= "Represented as * in the specification"

FT /tag= i
 FT /note= "Represented as * in the specification"

FT /tag= j
 FT /note= "Represented as * in the specification"

FT /tag= k
 FT /note= "Represented as * in the specification"

FT /tag= l
 FT /note= "Represented as * in the specification"

FT /tag= m
 FT /note= "Represented as * in the specification"

FT /tag= n
 FT /note= "Represented as * in the specification"

FT /tag= o
 FT /note= "Represented as * in the specification"

FT /tag= p
 FT /note= "Represented as * in the specification"

FT /tag= q
 FT /note= "Represented as * in the specification"


```

FT FT /tag= m
FT FT /note= "Represented as * in the specification"
FT FT 3941
FT FT misc_feature
FT FT /tag= o
FT FT /note= "Represented as * in the specification"
FT FT 3943
FT FT misc_feature
FT FT /tag= p
FT FT /note= "Represented as * in the specification"
FT FT 4361
FT FT misc_feature
FT FT /tag= g
FT FT /note= "Represented as * in the specification"

XX PN WO200133977-A1.
XX PD 17-MAY-2001.
XX PX
XX PE 06-NOV-2000; 2000WO-AU01362.
XX PR 05-NOV-1999; 99AU-0003875.
XX PA (META-) METABOLIC PHARM LTD.
XX PI Belyea CI, Ng FM, Vaughan P;
XX PS WPI; 2001-328876/34.
XX PT New organisms containing nucleic acid encoding a growth hormone
XX PT fragment which modulates lipid metabolism are useful to produce dietary
XX PT aids for obesity and in the meat production industry
XX PS Disclosure; Page 48-50; 54pp; English.
XX CC The invention relates to novel transgenic organisms useful in the
XX CC production of functional food and drink products for the treatment
XX CC or prevention of obesity via the regulation of lipid metabolism. The
XX CC organisms comprise a polynucleotide encoding a growth hormone fragment
XX CC capable of stimulating the activity of hormone-sensitive lipase (the key
XX CC enzyme in lipolysis) and inhibiting acetyl CoA carboxylase (the key
XX CC enzyme in lipogenesis). The growth hormone fragment preferably contains
XX CC at least the disulphide-bonded loop of a mammalian growth hormone (but is
XX CC not the full-length growth hormone) and is optionally linked to an
XX CC epitope tag or heterologous fusion protein partner. The transgenic
XX CC organism may be a microorganism used to produce a fermented product
XX CC (e.g., yeast), or an edible plant or animal or cell thereof. Food or
XX CC drink made using methods of the invention are used to modify fat/lean
XX CC ratio, lipid metabolism or food use in a mammal. In particular, the food
XX CC or drink products may be used to treat or prevent obesity, particularly
XX CC in humans, and may also be used to improve the fat/lean ration of
XX CC livestock raised for meat production. In the exemplification of the
XX CC invention, the human growth hormone (hGH) fragment analogue A096504 was
XX CC expressed in yeast, optionally fused to the FLAG epitope (AAB73625).
XX CC The present sequence is described as a DNA sequence from yeast in
XX CC the sequence listing, but is not further referred to in the
XX CC specification.
XX SQ Sequence 4590 BP; 661 A; 384 C; 127 G; 522 T; 2896 other;

Query Match 4.6%; Score 41.2; DB 22; Length 4590;
Best Local Similarity 10.0%; Pred. No. 0.54;
Matches 42; Conservative 210; Mismatches 169; Indels 0; Gaps 0;

```

```

QY 182 TGATGATATATGCGAAGACGACGATTTCATGTTAGAGCATCAGGCAATATCAT 241
DB 4037 NTHADRTTRRAANMBRNKNNBRSQMMKMMKWKYNBDBVRYTGGBDKMMNDYSMB 4096
QY 242 ATTCCAATGAGAGCCTTTCGCGAGAGAGGTCACACCCCGAAGAAAGATGACT 301
DB 4097 SHMMNDYSKDCMNDYSCMTATRCVRSANVASDMNDYSCMTATRCVRSANVASDMNDY 4156
QY 302 CCTGAAGTGAAGAGGCTTCTCCACCTCGATGAGAGAAAGCAATTCACAGA 361
DB 4157 MATRCVRSANVASDMNDYSCMTATRCVRSANVASDMNDYSCMTATRCVRSANVASDMNDY 4216
QY 362 GCGAAGAACTTCAGACCTGACATCTATCGAATTCAGAAATCTAATTAAGTGA 421
DB 4217 SCTMATRCVRSANVASDRSKMMKSSMGTSNRRMANSTMBARACRRNHHBBDIM 4276
QY 422 C 422
DB 4277 B 4277

RESULT 13
ABL32420
ID ABL32420 standard; DNA; 6079 BP.
XX ABL32420;
XX AC
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 393.
XX KW Human; immune system disease; cytosine methylation; antiaethmatic;
XX KW antiarteriosclerotic; antianaemic; cyrosatic; noctropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
XX KW antineumalator; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX KW gene; ds.
XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JUN-2002.
XX PF 02-JUL-2001; 2001WO-EP07537.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIC-) EPITENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation
XX PS Claim 1; SEQ ID NO 393; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX SQ Sequence 6079 BP; 1964 A; 94 C; 1241 G; 2780 T; 0 other;

```


Query Match 4.6%; Score 41.2; DB 24; Length 6079;
 Best Local Similarity 49.1%; Pred. No. 0.61;
 Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 584 TTGTGACATTTTATATCTCTCTCTAGTGGCTGATTTCTTACTTATCTCAATTTT 643
 DB 862 TTGTGATATTTTGTATGAGAGAGTTGATTAATTAATGATTAATTAATTAATTT 921
 QY 644 GGCACCTCAGACAAATTAGCCCATTAATTCACACCTGAGGGTGTGTTGAGA 703
 DB 922 TGTATTTAAATTAATTTAAATAATTTTAAATATGAGAAATTAATGTTGATGA 981
 QY 704 GGGATATGATTTTATGAGATGATATGCAATGTGCTTACGATTTGATGAATTT 763
 DB 982 TATGATTAATTTTAAAGATGGAATGAAATGATTAATTAAGATTTTAATTAAT 1041
 QY 764 CCCAGCTACTCTACAGTATTTTGTCATATTTGGAATG 805
 DB 1042 AAAAATGATATGATGAAATGATTAATTAATTAATTTGAATG 1083

RESULT 14
 AAF75571
 ID AAF75571 standard; DNA; 13015 BP.
 AC AAF75571;
 XX
 DT 09-MAY-2001 (first entry)
 DE Drosophila tetratricopeptide repeat protein 2 regulatory region.
 KW Animal model; polyglutamine tract; neurodegenerative disorder; HDJ1;
 KW heat shock protein 40; tetratricopeptide repeat protein 2; TPR2; stroke;
 KW myeloid leukemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma;
 KW cancer; ds.
 OS Drosophila sp.
 PN W0200112238-A1.
 XX
 PD 22-FEB-2001.
 XX
 PF 14-AUG-2000; 2000WO-US22496.
 XX
 PR 12-AUG-1999; 99US-0148933.
 PR 12-AUG-1999; 99US-0148934.
 PR 18-JAN-2000; 2000US-017047.
 PR 19-MAY-2000; 2000US-0205720.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Benzer S, Kazemi-Bafarjani P;
 XX
 PT WPI; 2001-147537/15.
 XX
 DR Identifying genes or other compounds that modulate polyglutamine
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease
 PT and Creutzfeldt-Jakob disease -
 XX
 PS Disclosure; Fig 17; 275pp; English.
 XX
 CC The present invention describes a method of screening for genes which
 CC modulate polyglutamine toxicity using animal models with polyglutamine
 CC sequences that cause toxicity in the animal. The model is preferably
 CC Drosophila, and toxic polyglutamine sequences include the human and
 CC Drosophila heat shock protein 40/HDJ1, tetratricopeptide repeat protein 2
 CC (TPR2) and myeloid leukemia factor 1 (MLF) genes. The model is useful
 CC for identifying treatments for neurodegenerative and proliferative
 CC disorders, including Alzheimer's disease, Parkinson's disease,
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar

CC ataxias, dentatorubropallidoluysian atrophy, Kennedy's disease, stroke,
 CC head trauma and cancer.
 XX
 SQ Sequence 13015 BP; 3746 A; 2643 C; 2515 G; 4111 T; 0 other;

Query Match 4.6%; Score 41; DB 22; Length 13015;
 Best Local Similarity 54.2%; Pred. No. 0.98;
 Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 552 TGAATGCTGAGACTTACGAGAAATGCGTGTGATTAATATCTCTCTAG 611
 DB 1103 TCATAGACTCGAATTAATTTAGAAATTAATTAATTAATTTCTTTCTTTT 1162
 QY 612 TTGGCTATTTCTTACTTATCTTCAATTTTGGACCTCAGAACAAATTAGCCATA 671
 DB 1163 TTGGTTGGTTTTTTTTTTTTTTTTTTTTTTCTTTCGCAACTTTTCGCGCTTC 1222
 QY 672 AATTCACACCTGAGGGGTGTTTGGAGAG 704
 DB 1223 ATTTTGACAGCCGAGAGAGTTCGTTGTTGAG 1255

RESULT 15
 ABL08614
 ID ABL08614 standard; cDNA; 21475 BP.
 AC ABL08614;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20324.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster.
 PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PERK) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR P-PSDB; ABB64511.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 20324; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutic and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB57737-AB812072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 21475 BP; 6025 A; 4524 C; 4394 G; 6532 T; 0 other;

Query Match 4.6%; Score 41; DB 23; Length 21475;
Best Local Similarity 54.2%; Pred. No. 1.2;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 552 TGAATGCTGAGACTTAGCAAAATGGCCCTGTTGTACATTATATCTCTCTCTAG 611
14133 TCATAGACTCGAATTATTTAGAAATTAATATATATTTGTTCTCTCTTTTTT 14192

QY 612 TTGGCTGATTTCTTACTTATCTTCATTTTGGCACCTCAGACAAATTAGCCATA 671
Db 14193 TTGGTTGGTTTCTTTTTTTTGTGTTTCTTCTGCAACACTTTCCGCTCTC 14252

QY 672 AATTCAACCTGAGGGGTGTGTTTGAGGAG 704
Db 14253 ATTTGACAGCCGAGGAGTGGGTGCTCAG 14285

Search completed: April 15, 2003, 18:57:46
Job time : 271.696 secs

Db 215 TGCATGATATGTCGAAACAGCAGCTTTCATATGTTAGGCCATCCAGCAATATCAAT 274
 QY 242 ATTCATAGGAGACCTTTTGGCCAGACAGTCAACCCCGAAGAAAGAAATGTAAT 301
 Db 275 ATTCATAGGAGACCTTTTGGCCAGACAGTCAACCCCGAAGAAAGAAATGTAAT 334
 QY 302 CCTGATGAGAGAGGAGGTTCTCCCACTCGATGAGAGAGAAACCAATTCAGAGA 361
 Db 335 CCTGATGAGAGAGGAGGTTCTCCCACTCGATGAGAGAGAAACCAATTCAGAGA 394
 QY 362 GCGAAGAACTTCCAGAGCTGCACTCAATCTATCGAAATTCAGAAATTTAAAGTAA 421
 Db 395 GCGAAGAACTTCCAGAGCTGCACTCAATCTATCGAAATTCAGAAATTTAAAGTAA 454
 QY 422 CTAAATATATGTCCTCCAAAGCTGAAACATGATGAGAAAGAAAGATTTGATGAGAA 481
 Db 455 CTAAATATATGTCCTCCAAAGCTGAAACATGATGAGAAAGAAAGATTTGATGAGAA 514
 QY 482 TAAAGAGCAGAGATGATTCATATAGCTCACTAAATTTATATATTTGATGATGAT 541
 Db 515 TAAAGAGCAGAGATGATTCATATAGCTCACTAAATTTATATATTTGATGATGAT 574
 QY 542 GTGAACCTCTGATGCTGAGACTTGAAGAAATGAGCTGTTGATCAATTTATATCTC 601
 Db 575 GTGAACCTCTGATGCTGAGACTTGAAGAAATGAGCTGTTGATCAATTTATATCTC 634
 QY 602 TTCTTTGATGCTGATGATTTTCTTATCTTATCTTATTTTGGAGACTTCAGAAACAA 661
 Db 635 TTCTTTGATGCTGATGATTTTCTTATCTTATCTTATTTTGGAGACTTCAGAAACAA 694
 QY 662 TTAGCCCATTAATTCACACCTGAGAGGTTGTTTGGAGAGGATATGATTTATGGA 721
 Db 695 TTAGCCCATTAATTCACACCTGAGAGGTTGTTTGGAGAGGATATGATTTATGGA 754
 QY 722 GAATGATATGCAATGTCCTTAACGATTTTATGAGAAAGTTTCCCAAGCTACTTCTACA 781
 Db 755 GAATGATATGCAATGTCCTTAACGATTTTATGATGAGAAAGTTTCCCAAGCTACTTCTACA 814
 QY 782 GTATTTTGGT 791
 Db 815 GTATTTTGGT 824

RESULT 2

US-09-962-436-468/C
 ; Sequence 468, Application US/09962436
 ; Patent No. US20020081301A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Soppet, Daniel
 ; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
 ; FILE REFERENCE: 669290-75
 ; CURRENT APPLICATION NUMBER: US/09/962,436
 ; CURRENT FILING DATE: 2001-09-25
 ; PRIOR APPLICATION NUMBER: US/60/235,082
 ; PRIOR FILING DATE: 2000-09-25
 ; PRIOR APPLICATION NUMBER: US/60/234,924
 ; PRIOR FILING DATE: 2000-09-25
 ; NUMBER OF SEQ ID NOS: 568
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 468
 ; LENGTH: 587
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: n=a,t,g or c
 US-09-962-436-468

Query Match 60.0%; Score 531.8; DB 10; Length 587;
 Best Local Similarity 97.1%; Pred. No. 6.5e-137;
 Matches 572; Conservative 0; Mismatches 14; Indels 3; Gaps 3;

QY 290 AAGATGTAATCTCTGAGATGAGAGAGGTTTCTCCCACTCGATGAGAGAGAG 349
 Db 587 AAGATGTAATCTCTCCCAAGTGAAGAGAGGTTG-TCTTCCCACTCGATGAGAGAGAG 529
 QY 350 CCAATTCAGAGAGAGAGAACTTCCAGAGCTGCACTCAATCTATCGAAATTCAGAAAT 409
 Db 528 CCAATTCAGAGAGAGAGAGAACTTCCAGAGCTGCACTCAATCTATCGAAATTCAGAAAT 470
 QY 410 ATTAAAGTGA-CTAAATATGTTCCCAAGCTGAAACGATGATGAGAGAGAGAGAT 468
 Db 469 ATTAAAGTGA-CTAAATATGTTCCCAAGCTGAAACGATGATGAGAGAGAGAGAT 410
 QY 469 TGATGTAAGAAATTAAGAGAGAGAGAGATGATGATGATGATGATGATGATGAT 528
 Db 409 TGATGTAAGAAATTAAGAGAGAGAGAGATGATGATGATGATGATGATGATGAT 350
 QY 529 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 588
 Db 349 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 290
 QY 589 ACATTTATATCT 648
 Db 289 ACATTTATATCT 230
 QY 649 CTCACAGAAATTAAGAT 708
 Db 229 CTCACAGAAATTAAGAT 170
 QY 709 ATGATTTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
 Db 169 ATGATTTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 110
 QY 769 GCTACTTCTTACAGATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 828
 Db 109 GCTACTTCTTACAGATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 50
 QY 829 ATTAGTCACTTAACCTTGTATGATGATGATGATGATGATGATGATGATGATGAT 877
 Db 49 ATTAGTCACTTAACCTTGTATGATGATGATGATGATGATGATGATGATGATGAT 1

RESULT 3

US-09-960-352-5216
 ; Sequence 5216, Application US/09960352
 ; Patent No. US20020137139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Wesley C.
 ; APPLICANT: Tao, Nengbing
 ; APPLICANT: Byatt, John C.
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 ; FILE REFERENCE: 16511.006/37-21(10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 5216
 ; LENGTH: 466
 ; TYPE: DNA
 ; ORGANISM: Bos taurus
 ; OTHER INFORMATION: Clone ID: 23-BOWMS1-014-Q1-E1-F3
 US-09-960-352-5216

Query Match 36.0%; Score 319; DB 10; Length 466;
 Best Local Similarity 83.9%; Pred. No. 3.3e-78;
 Matches 387; Conservative 0; Mismatches 65; Indels 9; Gaps 2;

QY 81 CCGGCGCGCCCTTCCACCTTCCAGAGAGCTTTGATTTTTCGATCTGGCTGGGACT 140
 Db 1 CCGGCGCTCCCTTCCACCTTCCAGAGAGCTTTGATTTTTCGATCTGGCTGGGACT 58
 QY 141 TCCCTTGGAGATTAACAATATACATAAGCAGGATTAAGCTGCAATGATATGCGAAG 200

Db	59	TTCCTCAGGCAATGAAACATGTGCATATAGCA	CGGATTAAGACCGGGTCAGTATGTCCAAAC	11.8
Qy	201	AGCCAGTTTCCAAITGTTAGAGCCATCCAGGCAATATCAATATTTCCAAATGGAGCCTTTTC		26.6
Db	119	AGCCAGTTTCCAAITGTCAAGCCCATCCATGCAAAATATCAATATTTCCCAATGGAGACCTTTTC		17.8
Qy	261	GGCCAGGAGCAGGTCAACCCCCCGAANAAGAAATGATCTCCGAAAGTGGAGGAGGGTG		32.0
Db	179	GGCCAGGAGCTGGGCACCCCTTAGAAGAAAAGATGTACTCCGAGATAGAGAGGGGTG		23.8
Qy	321	TTTCCTCCCACTCGGATGAGAGAAAGCAATTTCCAGAGCGAAGAACTTCCAGAGAC		38.0
Db	239	CTCCCTCCCACTTCCGACGATGAGAAAGACCAATTTCCAGAGGCTTAGAAACTTCCAGAAC		29.8
Qy	381	CTGCACTCAATCTATCGGAAATCCAGAAATTAAGGTGAATCTAAAAATATGTCCCAAG		44.0
Db	299	CTGCACTCAACTTATCGAGATTCGAACTTAAGAAAGTGAATCTGAATATGTCCCAAG		35.8
Qy	441	CTGAACATTAAGAGAAAGAAATTTGATGTGAGAAATTAAGAGGCGAAGATGGA		50.0
Db	359	CTGAACATTTGTGGAAAGAAAAACGCTTGATGTGAGAAATTTGAACAAGATTAGAGAAATGGA		41.8
Qy	501	TTT-----AATAGCTCACTAAATTTTATATATTTGTAT	53.4	
Db	419	TTCAAAACAAAATAGGCTCACTAAAGAAATTTTATATATGTGT	45.9	

RESULT 4
 US-09-070-927A-394/C
 Sequence 394, Application US/09070927A
 Patent NO. US20020120116A1
 GENERAL INFORMATION:
 APPLICANT: Charles A. Kunech
 Patrick J. Dillon
 Steven Barab
 TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 982
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/070,927A
 FILING DATE: 04-May-2000
 CLASSIFICATION: <Unknown> /
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/046,655
 FILING DATE: 1997-05-16
 APPLICATION NUMBER: 60/044,031
 FILING DATE: 1997-05-06
 APPLICATION NUMBER: 60/066,009
 FILING DATE: 1997-11-14
 ATTORNEY/AGENT INFORMATION:
 NAME: Kenley K. Hoover
 REGISTRATION NUMBER: 40,302
 REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 394:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 14141 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear

```

;          SEQUENCE DESCRIPTION: SEQ ID NO: 394:
US-09-070-927A-394

Query Match:          4.2%; Score 37.6; DB 10; Length 14141;
Best Local Similarity 52.6%; Pred. No. 13;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 730 TGGCAATGTCCTAAACGATTTTGATGAAAGTTTCCAGCTACTTCTACAGTATTTTG 789
Db 433 TCGGATTTGCTAAACAAACGACGACGATGTCGAAATTCACITTCACAAATGCATTTTT 4284

QX 790 GTCAATATTTTGGAATGCCCTTTAGTCTTTCACCTTTTAAATATATGTCACTAACTTTGA 849
Db 4283 TTCAATTATTTCACCTCACTTTTATCTGGTTACCTTCATTATATACAGAAATTAATTTTAA 4224

QY 850 TGAGTTCAATTAATATTTTGACTAAATGTAATTAAGT 885
Db 4223 TGCCTTCAATGAGAAAAACGAACTTCTTACGT 4188

```

```

RESULT 5
US-09-815-242-9307/c
: Sequence 9307, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zykkind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OR INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,078
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9307
: LENGTH: 2253
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(2253)
US-09-815-242-9307

Query Match 4.2%; Score 37.2; DB 10; Length 2253;
Best Local Similarity 56.6%; Pred. No. 5.7;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 562 AGACTCTAGCAGAAATGCGCTGTTGTACATTATATCTCTCTCTAGTTGGCTGTAT 621
Db 730 AGACTTCGCGGTTTGTAGCTTGCAGATCGATATTTCTCCAGCGTTGGACGATTTT 671
Qy 622 TTCTTACTTATCTTCAATTTTGGACCTCAGACAAATTAAGCCATAATTCAACAC 681

```

Db 670 TTCTCATTTTCTCTCAAAAAGCTGCTCCCAAAAAGACAGAAAAAGCATTAATTCAGAC 611
QY 682 CT 683
Db 610 TT 609

RESULT 6
US-09-938-842A-3453/C
; Sequence 3453, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3453
; LENGTH: 1875
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3453

Query Match 4.1%; Score 36.6; DB 9; Length 1875;
Best Local Similarity 55.0%; Pred. No. 7.4;
Matches 72; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 332 TCGATAGAGAGAAAGCCATTCCAGAGCAAACTTCCAGAGCTGCACTCAT 391
Db 637 TTGATTAAGCAGACGTAAATGATGATACCCAAAGTAAAGTGAATCTGCAACATA 578
QY 392 CTATCGAGAGAAAGCCATTCCAGAGCAAACTTCCAGAGCTGCACTCAT 451
Db 577 ATAAATGAATCAATGATTAATTAATGCTTTAAGACGCGCACTAAATTAATGATTA 518
QY 452 TAGAGAGAAAA 462
Db 517 AACATCAGA 507

RESULT 7
US-10-044-090-123
; Sequence 123, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044, 090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 123
; LENGTH: 3633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 892168.1
US-10-044-090-123

Query Match 4.1%; Score 36.6; DB 12; Length 3633;

Best Local Similarity 53.1%; Pred. No. 11;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 475 GAAGAAATAAGAGGAGAGATGATGATCAATGATCACTAAATTTATATATTTGTAT 534
Db 3121 GATTAAAAAAGAAAAAGCTATGACCAAAATGGCTTAAGATATTTTAAGGAAG 3180
QY 535 GATGATTGTGACCTCTGATGATGCTGAGACTAGAGCAAAATGGCTGTTGTACATT 594
Db 3181 AAGATATTAAGCAACAGTTATACATGAGACTATGTTGACTCTTTCTTTACCTT 3240
QY 595 ATATCTCTCTCTGATGCTGTAT 621
Db 3241 AAGCTTAAGATTTCTTTTATGATAT 3267

RESULT 8
US-09-070-927A-89
; Sequence 89, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; APPLICANT: Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070, 927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 89:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20072 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-070-927A-89

Query Match 4.1%; Score 36.6; DB 10; Length 20072;
Best Local Similarity 48.8%; Pred. No. 30;
Matches 99; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 670 TAAATTCACACCTGAGGGTGTGTTTGAAGAGGATATGATTTATGAGAGATGATA 729
Db 17319 TAAATGTGGCCCTGAACGTTGTTTGAATGATACCATCTGTATTAATTTAAAT 17378

```

1 LENGTH: 395
2 TYPE: DNA
3 ORGANISM: Homo sapiens
4 FEATURE:
5 OTHER INFORMATION: MAP TO AL033539.17
6 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
7 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4

```

RESULT 11
US-10-025-380-275/c
; Sequence 275, Application US/10025380

? APPLICANT: Xu, Jiangchun
 ? APPLICANT: Lodes, Michael J.
 ? APPLICANT: Secrist, Heather
 ? APPLICANT: Benson, Darin R.
 ? APPLICANT: Mesgher, Madeleine Joy
 ? APPLICANT: Stolk, John A.
 ? APPLICANT: Wang, Tongtong
 ? APPLICANT: Jiang, Yugu

OTHER INFORMATION
US-09-833-263-275

4.1%; Score 36.2; DB 10; Length 528;

Best Local Similarity 48.7%; Pred. No. 4.5;
Matches 128; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

QY 339 AGGAGAAGAGCCCAATTCAGAGCGAAGAACTTCAGAGCCTGAGCATCTATCCG 398
DB 390 AGAGACGACAGCAATTTGGCGCAATTAATMAAGACCCAGAGCTCCAAATTTTCCAAAGTG 331
QY 399 --AAATCCGAATATTAAAGTGAACATAATATGTCCCAAGCTGAACAGTAGAGA 456
DB 330 CAGATTATGGAATAGTTCAGATTCAGATTATTAAGTAGTTCCTGAAATGACTGAGATTTGA 271
QY 457 AGAAAAAGAGATTGATGTGAAGAAATTAAGAGCGAAGATGATTCATTAAGCTCACTAA 516
DB 270 AGAAAAAGATTCAGATTCAGATTCAGCTTAAAAAGAAAACTTTGTTAAAGTATTCACCTGA 211
QY 517 AATTTATATATTTGTATGATGATGTGAACCTCCGTAATGCTGAGACTTCAGAGAAA 576
DB 210 AATCAGAGATATTGGGTATTTATTAACAATCATTCGAAAGATGAGAGCTACATTTCA 151
QY 577 TGGCTGTTGTACATTATATC 599
DB 150 TAATTTGAGGAAAAATTTCTTAC 128

RESULT 14

US-09-535-459-2163
Sequence 2163, Application US/09535459
Publication No. US20030040615A1
GENERAL INFORMATION:
APPLICANT: Seilheimer, Jeffrey J.
APPLICANT: Deleage, Angelo M.
APPLICANT: Stuart, Susan G.
APPLICANT: Mullaly, Sara J.
APPLICANT: Naughton, Rebecca E.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING ELECTRON TRANSPORT MOLECULE
FILE REFERENCE: PD-1014 CIP
CURRENT APPLICATION NUMBER: US/09/535,459
Prior Filing Date: 2000-03-24
Prior Application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 2170
SOFTWARE: PERL Program
SEQ ID NO 2163
LENGTH: 548
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030040615A1 hu00992223
US-09-535-459-2163

Query Match 4.1%; Score 36.2; DB 9; Length 548;
Best Local Similarity 48.7%; Pred. No. 4.6;
Matches 128; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

QY 339 AGGAGAAGAGCCCAATTCAGAGCGAAGAACTTCAGAGCCTGAGCATCTATCCG 398
DB 203 AGAGACGACAGCAATTTGGCGCAATTAATMAAGACCCAGAGCTCCAAATTTTCCAAAGTG 262
QY 399 --AAATCCGAATATTAAAGTGAACATAATATGTCCCAAGCTGAACAGTAGAGA 456
DB 263 CAGATTATGGAATAGTTCAGATTCAGATTATTAAGTAGTTCCTGAAATGACTGAGATTTGA 322
QY 457 AGAAAAAGAGATTGATGTGAAGAAATTAAGAGCGAAGATGATTCATTAAGCTCACTAA 516
DB 323 AGAAAAAGATTCAGATTCAGATTCAGCTTAAAAAGAAAACTTTGTTAAAGTATTCACCTGA 382
QY 517 AATTTATATATTTGTATGATGATGTGAACCTCCGTAATGCTGAGACTTCAGAGAAA 576
DB 383 AATCAGAGATATTGGGTATTTATTAACAATCATTCGAAAGATGAGAGCTACATTTCA 442
QY 577 TGGCTGTTGTACATTATATC 599

DB 443 TAATTTGAGGAAAAATTTCTTAC 465

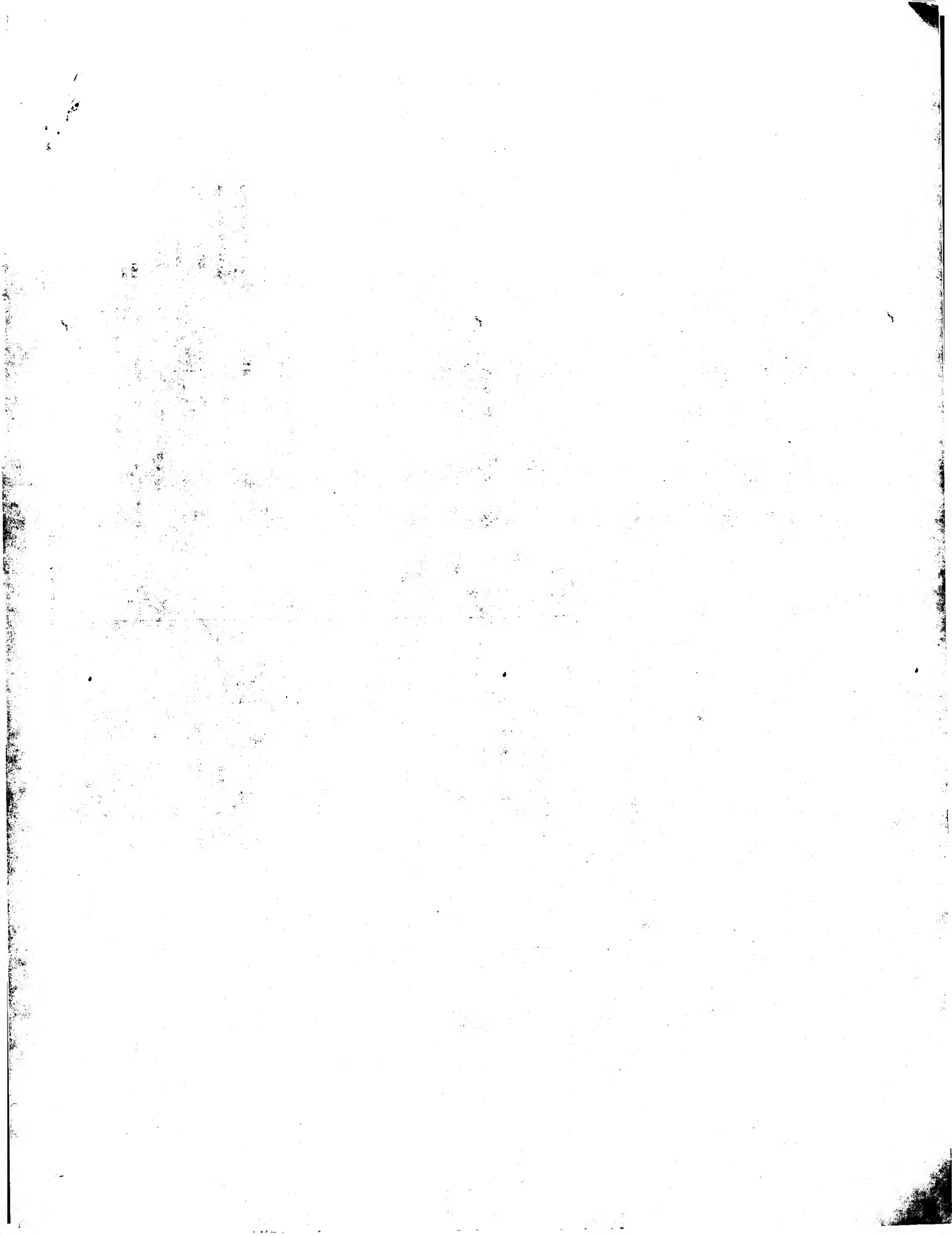
RESULT 15

US-09-969-347-175
Sequence 175, Application US/09969347
Patent No. US20020115085A1
GENERAL INFORMATION:
APPLICANT: Ebner, Reinhard
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signa
FILE REFERENCE: 689290-69
CURRENT APPLICATION NUMBER: US/09/969,347
Prior Filing Date: 2001-10-02
Prior Application Number: US/60/237,598
Prior Filing Date: 2000-10-03
Prior Application Number: US/60/237,604
Prior Filing Date: 2000-10-03
NUMBER OF SEQ ID NOS: 318
SOFTWARE: PatentIn version 3.0
SEQ ID NO 175
LENGTH: 1266
TYPE: DNA
ORGANISM: Homo sapiens
US-09-969-347-175

Query Match 4.1%; Score 36.2; DB 10; Length 1266;
Best Local Similarity 48.7%; Pred. No. 7.6;
Matches 128; Conservative 0; Mismatches 133; Indels 2; Gaps 1;

QY 339 AGGAGAAGAGCCCAATTCAGAGCGAAGAACTTCAGAGCCTGAGCATCTATCCG 398
DB 872 AGAGACGACAGCAATTTGGCGCAATTAATMAAGACCCAGAGCTCCAAATTTTCCAAAGTG 931
QY 399 --AAATCCGAATATTAAAGTGAACATAATATGTCCCAAGCTGAACAGTAGAGA 456
DB 932 CAGATTATGGAATAGTTCAGATTCAGATTATTAAGTAGTTCCTGAAATGACTGAGATTTGA 991
QY 457 AGAAAAAGAGATTGATGTGAAGAAATTAAGAGCGAAGATGATTCATTAAGCTCACTAA 516
DB 992 AGAAAAAGATTCAGATTCAGATTCAGCTTAAAAAGAAAACTTTGTTAAAGTATTCACCTGA 1051
QY 517 AATTTATATATTTGTATGATGATGTGAACCTCCGTAATGCTGAGACTTCAGAGAAA 576
DB 1052 AATCAGAGATATTGGGTATTTATTAACAATCATTCGAAAGATGAGAGCTACATTTCA 1111
QY 577 TGGCTGTTGTACATTATATC 599
DB 1112 TAATTTGAGGAAAAATTTCTTAC 1134

Search completed: April 16, 2003, 01:13:08
Job time : 123.63 secs



GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 ; Search time 2009.65 seconds
(without alignments)

12845.099 Million cell updates/sec

Title: US-09-647-019-3

Perfect score: 887

Sequence: 1 ggtctcctaaccgagagag.....tgactaaatgtaaatgta 887

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GeneBdb: 1: gb_ba: 2: gb_hcg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pac: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_scs: 12: gb_ey: 13: gb_un: 14: gb_vl: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pac: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_scs: 28: em_un: 29: em_vl: 30: em_hcg_hum: 31: em_hcg_inv: 32: em_hcg_other: 33: em_hcg_mus: 34: em_hcg_pln: 35: em_hcg_rtd: 36: em_hcg_mam: 37: em_hcg_vit: 38: em_sy: 39: em_hcgo_hum: 40: em_hcgo_mus: 41: em_hcgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	886	99.9	886	9	AF129505 Homo sapi
2	885	99.8	885	9	HS250584 Homo sapi
3	884.4	99.7	886	6	AX322774 Sequence
4	808.4	91.1	835	6	BC005948 Homo sapi
5	531.8	60.0	587	6	AX332418 Sequence
6	475.6	53.6	892	10	AF364071 Sequence
7	461	52.0	936	10	AF364070 Sequence
8	454.4	51.2	943	10	MM0245772 Homo sapi
9	424	47.8	36503	9	U73508 Homo sapi
10	424	47.8	124186	9	AL772332 Human DNA
11	389.8	43.9	787	10	AY026524 Sequence
12	290	32.7	290	6	AX322783 Sequence
13	221.4	25.0	228031	2	AL732396 Mus muscu
14	210.6	23.7	188670	2	AC127606 Rattus no
15	207.4	23.4	82586	2	AC096040 Rattus no
16	172.4	19.4	47440	9	U73509 Homo sapi
17	172.4	19.4	150319	2	AL772370 Homo sapi
18	141.6	16.0	923	5	AF343894 Xenopus l
19	96.6	10.9	239667	2	AL731811 Mus muscu
20	62.8	7.1	82586	2	AC096040 Rattus no
21	60	6.8	7218	6	I66494 Sequence
22	49.2	5.5	203251	2	AC124570 Mus muscu
23	48.8	5.5	2781	3	AY044085 Dictyoste
24	46.8	5.3	146383	2	AC116367 Oryza sat
25	46.2	5.2	1141	6	AX083744 Sequence
26	46	5.2	49843	3	AF288092 Neisseria
27	45.8	5.2	53440	2	AC117078 Dictyoste
28	45.2	5.1	7622	3	AF112367 Plasmodu
29	44.8	5.1	171187	2	AC116960 Dictyoste
30	44.6	5.0	115360	2	AC105578 Rattus no
31	44.4	5.0	156337	2	AC009938 Homo sapi
32	44.2	5.0	186634	2	AC100722 Mus muscu
33	44.2	5.0	190273	2	AL844901 Mus muscu
34	44.2	5.0	142396	9	AL157407 Human DNA
35	43.8	4.9	1141	6	AX083744 Sequence
36	43.6	4.9	104014	2	AC116921 Dictyoste
37	43.6	4.9	150552	2	AC121952 Mus muscu
38	43.4	4.9	55470	2	AC116989 Dictyoste
39	43.2	4.9	34087	3	CEH02112 Caenorhabdi
40	43.2	4.9	197900	9	AC002531 Homo sapi
41	43	4.8	6115	6	AX346702 Sequence
42	43	4.8	27046	3	AC115581 Dictyoste
43	43	4.8	93491	2	AC116967 Dictyoste
44	42.8	4.8	13604	3	NM417719 Nectator a
45	42.8	4.8	80250	5	AL691516 Zebrafish

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
AF129505	AF129505	Homo sapiens small muscular protein (SMPX) mRNA, complete cds.	AF129505	AF129505.1	GI:6625646	Homo sapiens.	Homo sapiens	1 (bases 1 to 886)	Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.	X-chromosomal human gene (SMPX) encoding a small muscular protein
AF129505	AF129505	886 bp mRNA, linear, PRI 22-DEC-1999	AF129505	AF129505.1	GI:6625646	Homo sapiens.	Homo sapiens	1 (bases 1 to 886)	Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.	X-chromosomal human gene (SMPX) encoding a small muscular protein

JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
 MEDLINE 2005879
 PUBMED 10598820
 REFERENCE 2 (bases 1 to 886)
 AUTHORS Patzak, D.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-1999) Molecular Human Genetics, Institut for Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany

FEATURES
 source
 1..886
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xq22.1-22.2"
 <1..>886
 /gene="SMPX"
 <1..183
 /gene="SMPX"
 184..450
 /gene="SMPX"
 /codon_start=1
 /evidence=not experimental
 /product="small muscular protein"
 /protein_id="AAFI9343.1"
 /db_xref="GI:6625647"
 /translation="NMMSKOPVSNVRAIOANINIPMGARPGAGOPRRKECTPEVEE
 GVPTSDERKRPFGAKLPGRVNVLSIIONIKSELKYPRKAEQ"

misc_feature
 190
 /gene="SMPX"
 /note="alternate position for initiation methionine"
 451..>886
 /gene="SMPX"
 857..862
 /gene="SMPX"

3' UTR
 poly_a_signal
 857..862
 /gene="SMPX"

BASE COUNT 279 a 172 c 191 g 244 t

ORIGIN

Query Match 99.9%; Score 886; DB 9; Length 886;
 Best Local Similarity 100.0%; Pred. No. 1.7e-208; Mismatches 0; Gaps 0;
 Matches 886; Conservative 0; Indels 0;

QY 2 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGAATTGA 61
 DB 1 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGAATTGA 60
 QY 62 GATCGAGCTCAGAGAGACCGGGGCCCCCTTCCACTTCCAGAGAGCTTGTATTCTTG 121
 DB 61 GATCGAGCTCAGAGAGACCGGGGCCCCCTTCCACTTCCAGAGAGCTTGTATTCTTG 120
 QY 122 CATCTGGCTGCTGGGACTTCCCTTAGGCACTAAACAATCATTAACGAGATTAAGAC 181
 DB 121 CATCTGGCTGCTGGGACTTCCCTTAGGCACTAAACAATCATTAACGAGATTAAGAC 180
 QY 182 TGCATGATATGTGGAACACGCCAGTTTCCATGTTAGAGCATTCAGCAAAATATCAAT 241
 DB 181 TGCATGATATGTGGAACACGCCAGTTTCCATGTTAGAGCATTCAGCAAAATATCAAT 240
 QY 242 ATTCGAATGGAGGCTTTGGCCAGAGAGAGTCAACCCCCAGAGAGAAAGAAATGTA 301
 DB 241 ATTCGAATGGAGGCTTTGGCCAGAGAGAGTCAACCCCCAGAGAGAAAGAAATGTA 300
 QY 302 CCTGAAGTGGAGAGGAGGTTCTCCCACTTCGATGAGAGAGAAAGCAATTCAGAGA 361
 DB 301 CCTGAAGTGGAGAGGAGGTTCTCCCACTTCGATGAGAGAGAAAGCAATTCAGAGA 360
 QY 362 GCGAAGAACTTCCAGAGCTGCACTCATTCGAAATCCAGAAATATTTAAAGTGA 421
 DB 361 GCGAAGAACTTCCAGAGCTGCACTCATTCGAAATCCAGAAATATTTAAAGTGA 420
 QY 422 CTAAATATGTGCTCCCAAGCTGAAACAGTAGTAGAGAGAAAGAGATGATGTAGAGAA 481
 DB 421 CTAAATATGTGCTCCCAAGCTGAAACAGTAGTAGAGAGAAAGAGATGATGTAGAGAA 480

QY 482 TAAAGAGGAGAGATGATTCATAGCTCACTAAATTTATATATTGATGATGAT 541
 DB 481 TAAAGAGGAGAGATGATTCATAGCTCACTAAATTTATATATTGATGATGAT 540
 QY 542 GTGAACCTCCTGAATGCTGAGACTTAGAGAGAAATGAGCTGTTTGATATATATCTC 601
 DB 541 GTGAACCTCCTGAATGCTGAGACTTAGAGAGAAATGAGCTGTTTGATATATATCTC 600
 QY 602 TTCTCTTAGTGGCTGTATTTCTTACTTATCTTCAATTTTGGACCTCAAGAAACAA 661
 DB 601 TTCTCTTAGTGGCTGTATTTCTTACTTATCTTCAATTTTGGACCTCAAGAAACAA 660
 QY 662 TTAGCCATAAATCAACACCTGGAGGAGTGCTTTGAGAGAGGATATATTTATGGA 721
 DB 661 TTAGCCATAAATCAACACCTGGAGGAGTGCTTTGAGAGAGGATATATTTATGGA 720
 QY 722 GAATGATATGGCAATGTCCTTAACGATTTTGATGAGAAAGTTCCAGACTTCTCTACA 781
 DB 721 GAATGATATGGCAATGTCCTTAACGATTTTGATGAGAAAGTTCCAGACTTCTCTACA 780
 QY 782 GTATTTGGTCAATATTTGGAATGCTTTAGTCTTCACTTTTAATATATGCTAA 841
 DB 781 GTATTTGGTCAATATTTGGAATGCTTTAGTCTTCACTTTTAATATATGCTAA 840
 QY 842 ACTTGTATGAGTCAATATAATTTGACTTAATGTAAATGGA 887
 DB 841 ACTTGTATGAGTCAATATAATTTGACTTAATGTAAATGGA 886

RESULT 2
 HSA250584 885 bp mRNA linear PRI 12-APR-2001
 LOCUS HSA250584
 DEFINITION Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
 ACCESSION AJ250584
 VERSION AJ250584.1 GI:10178976
 KEYWORDS Strux gene; stretch responsive muscle (X-chromosome).
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 885)
 Kemp, T.J., Sadusky, T.J., Simon, M., Brown, R., Eastwood, M.,
 Sassoon, D.A. and Coulton, G.R.
 Identification of a novel stretch-responsive skeletal muscle gene
 (Smpx)
 Genomics 72 (3), 260-271 (2001)
 JOURNAL
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2 (bases 1 to 885)
 Kemp, T.J.
 Direct Submission
 Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial
 College of Science Technology and Medicine, SAF Bldg, Level 2,
 Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
 COMMENT Related sequences: AJ245772, U73508 to U73509.
 FEATURES
 source
 1..885
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="skeletal muscle"
 1..885
 /gene="Smpx"
 1..183
 /gene="Smpx"
 184..450
 /gene="Smpx"
 /codon_start=1
 /product="stretch responsive muscle (X-chromosome)"
 /protein_id="CAC08492.1"
 /db_xref="GI:10178977"
 /translation="NMMSKOPVSNVRAIOANINIPMGARPGAGOPRRKECTPEVEE
 GVPTSDERKRPFGAKLPGRVNVLSIIONIKSELKYPRKAEQ"

3'UTR 451.885
/gene="Srmx"
polyA_signal 857.862
/gene="Srmx"
BASE COUNT 278 a 172 c 191 g 244 t
ORIGIN

Query Match 99.8%; Score 885; DB 9; Length 885;
Best Local Similarity 100.0%; Pred. No. 3e-208;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTCTCAATACCGGAGAGGACAGACGCTATTTCAGCCACATGAAAAGCATCGGAATTGA 61
DB 1 GTTCTCAATACCGGAGAGGACAGACGCTATTTCAGCCACATGAAAAGCATCGGAATTGA 60
QY 62 GATCGAGCTCAGAGGACACCGGGGCGCCCTTCCACCTTCCAGAGACCTTGTATCTTG 121
DB 61 GATCGAGCTCAGAGGACACCGGGGCGCCCTTCCACCTTCCAGAGACCTTGTATCTTG 120
QY 122 CATCTGCTGCTGGGACTTCCCTTAGGCACTAAACAATACATTAAGACAGGATAAGAC 181
DB 121 CATCTGCTGCTGGGACTTCCCTTAGGCACTAAACAATACATTAAGACAGGATAAGAC 180
QY 182 TGCATGAAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 241
DB 181 TGCATGAAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 240
QY 242 ATTCGCAATGGAGGCTTTCCGGCCAGAGAGCTCAACCCCGAAGAAAAGAAATGTAAT 301
DB 241 ATTCGCAATGGAGGCTTTCCGGCCAGAGAGCTCAACCCCGAAGAAAAGAAATGTAAT 300
QY 302 CCTGAATGAGAGAGGCTTCTCCCACTCGATGAGAGAGAAAGCAATTCAGAGA 361
DB 301 CCTGAATGAGAGAGGCTTCTCCCACTCGATGAGAGAGAAAGCAATTCAGAGA 360
QY 362 GCGAAGAACTTCCAGAGACCTGCACTCAATCTATCGAAATCCAGAAATTTAAAGTAA 421
DB 361 GCGAAGAACTTCCAGAGACCTGCACTCAATCTATCGAAATCCAGAAATTTAAAGTAA 420
QY 422 CTAATAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 481
DB 421 CTAATAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 480
QY 482 TAAAGAGGCAAGAGATGATTCATATAGCTCACTAAATTTATATTTGATGATGATT 541
DB 481 TAAAGAGGCAAGAGATGATTCATATAGCTCACTAAATTTATATTTGATGATGATT 540
QY 542 GTGAACCTCTGAAATGCTGAGACTGACAGAAATGAGCTGTTGTACATTTATATCTC 601
DB 541 GTGAACCTCTGAAATGCTGAGACTGACAGAAATGAGCTGTTGTATATTTATATCTC 600
QY 602 TTCCTTCTAGTGGCTGTATTTCTTACTTTATCTTCAATTTTGGCACCCTCAAGAACAA 661
DB 601 TTCCTTCTAGTGGCTGTATTTCTTACTTTATCTTCAATTTTGGCACCCTCAAGAACAA 660
QY 662 TTAGCCCATTAATTCACACCTGAGAGGCTGTTTGAAGAGGAGATATGATTTATGGA 721
DB 661 TTAGCCCATTAATTCACACCTGAGAGGCTGTTTGAAGAGGAGATATGATTTATGGA 720
QY 722 GAATGATATGCAATGTCCTAACGATTTTGTATGATAAAAGTTTCCCAAGCTACTTCTACA 781
DB 721 GAATGATATGCAATGTCCTAACGATTTTGTATGATAAAAGTTTCCCAAGCTACTTCTACA 780
QY 782 GATATTTGGTCAATATTTGGAATGCGTTTATGTTTCACTTTTAAATTTATGTCATTA 841
DB 781 GATATTTGGTCAATATTTGGAATGCGTTTATGTTTCACTTTTAAATTTATGTCATTA 840
QY 842 ACTTGTATGAGTTCAATTAATATTTGACTAAATGTAATATG 886
DB 841 ACTTGTATGAGTTCAATTAATATTTGACTAAATGTAATATG 885

RESULT 3

AX322774
LOCUS AX322774 886 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 18 from Patent WO0192567.
ACCESSION AX322774
VERSION AX322774.1 GI:18093754
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1
AUTHORS Bunk, D., Reuner, B., Beck, J. and Henkel, T.
TITLE Novel target genes for diseases of the heart
JOURNAL Patent: WO 0192567-A 18 06-DEC-2001;
Medicine AG (DE)
FEATURES
source location/Qualifiers
1.886
BASE COUNT 278 a 172 c 191 g 245 t
ORIGIN

Query Match 99.7%; Score 884.4; DB 6; Length 886;
Best Local Similarity 99.9%; Pred. No. 4.2e-208;
Matches 885; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTTCTCAATACCGGAGAGGACAGACGCTATTTCAGCCACATGAAAAGCATCGGAATTGA 61
DB 1 GTTCTCAATACCGGAGAGGACAGACGCTATTTCAGCCACATGAAAAGCATCGGAATTGA 60
QY 62 GATCGAGCTCAGAGGACACCGGGGCGCCCTTCCACCTTCCAGAGACCTTGTATCTTG 121
DB 61 GATCGAGCTCAGAGGACACCGGGGCGCCCTTCCACCTTCCAGAGACCTTGTATCTTG 120
QY 122 CATCTGCTGCTGGGACTTCCCTTAGGCACTAAACAATACATTAAGACAGGATAAGAC 181
DB 121 CATCTGCTGCTGGGACTTCCCTTAGGCACTAAACAATACATTAAGACAGGATAAGAC 180
QY 182 TGCATGAAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 241
DB 181 TGCATGAAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 240
QY 242 ATTCGCAATGGAGGCTTTCCGGCCAGAGAGCTCAACCCCGAAGAAAAGAAATGTAAT 301
DB 241 ATTCGCAATGGAGGCTTTCCGGCCAGAGAGCTCAACCCCGAAGAAAAGAAATGTAAT 300
QY 302 CCTGAATGAGAGAGGCTTCTCCCACTCGATGAGAGAGAAAGCAATTCAGAGA 361
DB 301 CCTGAATGAGAGAGGCTTCTCCCACTCGATGAGAGAGAAAGCAATTCAGAGA 360
QY 362 GCGAAGAACTTCCAGAGACCTGCACTCAATCTATCGAAATCCAGAAATTTAAAGTAA 421
DB 361 GCGAAGAACTTCCAGAGACCTGCACTCAATCTATCGAAATCCAGAAATTTAAAGTAA 420
QY 422 CTAATAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 481
DB 421 CTAATAATATGTCGAAACAGCCAGTTTCCAAATGTAAGCCATCCAGGCAATATCAAT 480
QY 482 TAAAGAGGCAAGAGATGATTCATATAGCTCACTAAATTTATATTTGATGATGATT 541
DB 481 TAAAGAGGCAAGAGATGATTCATATAGCTCACTAAATTTATATTTGATGATGATT 540
QY 542 GTGAACCTCTGAAATGCTGAGACTGACAGAAATGAGCTGTTGTACATTTATATCTC 601
DB 541 GTGAACCTCTGAAATGCTGAGACTGACAGAAATGAGCTGTTGTATATTTATATCTC 600
QY 602 TTCCTTCTAGTGGCTGTATTTCTTACTTTATCTTCAATTTTGGCACCCTCAAGAACAA 661
DB 601 TTCCTTCTAGTGGCTGTATTTCTTACTTTATCTTCAATTTTGGCACCCTCAAGAACAA 660
QY 662 TTAGCCCATTAATTCACACCTGAGAGGCTGTTTGAAGAGGAGATATGATTTATGGA 721
DB 661 TTAGCCCATTAATTCACACCTGAGAGGCTGTTTGAAGAGGAGATATGATTTATGGA 720

QY	722	GAATATATATGGCAATAGTCCCTAAACGATTTTATGATGAAAGTTTCCCAAGCTACTCTCTACA	781
QY <td>721</td> <td>GAATATATATGGCAATAGTCCCTAAACGATTTTATGATGAAAGTTTCCCAAGCTACTCTCTACA</td> <td>780</td>	721	GAATATATATGGCAATAGTCCCTAAACGATTTTATGATGAAAGTTTCCCAAGCTACTCTCTACA	780
Db <td>782</td> <td>GAATTTTGGTCAATATTTGGAAATGGGTTTNGTCTTCAACCTTTAAATATATGCTACTAA</td> <td>841</td>	782	GAATTTTGGTCAATATTTGGAAATGGGTTTNGTCTTCAACCTTTAAATATATGCTACTAA	841
QY <td>781</td> <td>GAATTTTGGTCAATATTTGGAAATGGGTTTNGTCTTCAACCTTTAAATATATGCTACTAA</td> <td>840</td>	781	GAATTTTGGTCAATATTTGGAAATGGGTTTNGTCTTCAACCTTTAAATATATGCTACTAA	840
Db <td>842</td> <td>ACTTGTATATGAGTTCAATATAATATTTGACTCTAAATATGTAATAATGTGA</td> <td>887</td>	842	ACTTGTATATGAGTTCAATATAATATTTGACTCTAAATATGTAATAATGTGA	887
QY <td>841</td> <td>ACTTGTATATGAGTTCAATATAATATTTGACTCTAAATATGTAATAATGTGA</td> <td>886</td>	841	ACTTGTATATGAGTTCAATATAATATTTGACTCTAAATATGTAATAATGTGA	886

RESULT 4	LOCUS	DEFINITION	IMAGE:424501	835 bp	mRNA	linear	PRI 12-JUL-2001
BC005948	BC005948	Homo sapiens, small muscle protein, X-linked, clone MGC:14584	IMAGE:424501, mRNA, complete cds.				

VERSION	EC005948.1	GI:13543590
KEYWORDS	MGC.	
SOURCE	Homo sapiens.	

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 835)
Strausberg,R.
Direct Submission
Submitted (02-APR-2001) National Institutes of Health, Mammalian

REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk

Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (HLNLT)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-sbnc.stanford.edu>
Contact: (Dickson, Mark) mcc@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
J. M.

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6625546, through the I.M.A.G.E. Consortium/BLAT at: <http://image.lln.gov> Series: IRAL Plate: 21 Row: a Column: 4

FEATURES	Location/Qualifiers
source	1. .835

```

/organism="Homo sapiens"
/db xref="locusID:23676"
/db xref="taxon:9606"
/clone="MGC:14584 IMAGE:4246501"
/tissue_type="Skeletal Muscle"
/clone_lib="Nih_MGC_81"
/lab_host="DH10B"
/host="Vector: pDNR-LIB"
113..379

```

BASE COUNT
ORIGIN

Query Match	91.1%	Score 808.4;	DB 9;	Length 835;
Best Local Similarity	99.9%	Pred. No. 2.6e-189;		
Matches, 809; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0

OY	73	AGAGGACACCGGGGCCCTTCCACCTTCCAAAGAGCTTGTATCTTGCATCTGGCTGC	132
Db	1	AGAGGACACCGGGGCCCTTCCACCTTCCAAAGAGCTTGTATCTTGCATCTGGCTGC	60
OY	133	CTGGGACTTCCCTTAGGACGATMAACAATACTAAAGACGGATVAGACTGCGATGAATAT	192
Db	61	CTGGGACTTCCCTTAGGACGATMAACAATACTAAAGACGGATVAGACTGCGATGAATAT	120
OY	193	GTCCGAAACAGCAGTTCATATGTATGAGCCATCCAGGCAATATCAATATTCGAATGGG	252
Db	121	GTCCGAAACAGCAGTTCATATGTATGAGCCATCCAGGCAATATCAATATTCGAATGGG	180
OY	253	AGCCTTTGGGCGAGGAGGAGGTCAACCCCGAGAAAGAAAGATGTACTCTCGAAGTGA	312
Db	181	AGCCTTTGGGCGAGGAGGAGGTCAACCCCGAGAAAGAAAGATGTACTCTCGAAGTGA	240
OY	313	GGAGGGTGTCTCTCCACCTCGGATGAGAGAGAGAACCCOATTCGAGAGCGGAAGAACT	372
Db	241	GGAGGGTGTCTCTCCACCTCGGATGAGAGAGAGAACCCOATTCGAGAGCGGAAGAACT	300
OY	373	TCCAGGACCTCGAGCATCTATGCGAATCCAGAAATTTAAATGAACTGAACTGAAATATGT	432
Db	301	TCCAGGACCTCGAGCATCTATGCGAATCCAGAAATTTAAATGAACTGAAATATGT	360
OY	433	CCCCAAGCTGAACAGTATGAGAGAGAGAGAGAGATGATGAGAGAAATAAAGAGGAG	492
Db	361	CCCCAAGCTGAACAGTATGAGAGAGAGAGAGAGATGATGAGAGAAATAAAGAGGAG	420
OY	493	AAGATGATTCATATGCTCACTAAATTTTATATTTGTATGATGATTTGAACTCTCT	552
Db	421	AAGATGATTCATATGCTCACTAAATTTTATATTTGTATGATGATTTGAACTCTCTCT	480
OY	553	GAATGCTGAGAGCTGTACAGAAATGGCTGTGTGATCATTTATCTCTTCCCTCTAGT	612
Db	481	GAATGCTGAGAGCTGTACAGAAATGGCTGTGTGATCATTTATCTCTTCCCTCTAGT	540
OY	613	TGGCTGATTTCTTACTTTATCTTCATTTTGGCAGCTCAGAGAACAAATTAGCCCATPA	672
Db	541	TGGCTGATTTCTTACTTTATCTTCATTTTGGCAGCTCAGAGAACAAATTAGCCCATPA	600
OY	673	ATTCAACACTGAGAGGGGTGTGTTTGGAGGGGATATGATTTTATGAGAAATGATATGG	732
Db	601	ATTCAACACTGAGAGGGGTGTGTTTGGAGGGGATATGATTTTATGAGAAATGATATGG	660
OY	733	CAATGTGCTTACAGATTTTGTATGAGAAAGTTTCCCAAGCTACTTCTTACAGATTTTGGTC	792
Db	661	CAATGTGCTTACAGATTTTGTATGAGAAAGTTTCCCAAGCTACTTCTTACAGATTTTGGTC	720
OY	793	AATATTTGGAATGCGTTTATGTTCTTCACTTTTAAATATATGCTAAACTTTGTATGA	852
Db	721	AATATTTGGAATGCGTTTATGTTCTTCACTTTTAAATATATGCTAAACTTTGTATGA	780
OY	853	GTTCAAAATTAATATTTGACTAAATGTAAA 882	
Db	781	GTTCAAAATTAATATTTGACTAAATGTAAA 810	

RESULT 5				
AX332418/c	AX332418	587 bp	DNA	linear
LOCUS	Sequence 2927 from Patent WO0194629.			PAT 09-JAN-2002
DEFINITION	AX332418			
ACCESSION	AX332418			
VERSION	AX332418.1	GI:18123052		
KEYWORDS	.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G., Harrisgan,S., Soppet,D.R. and Weaver,Z. Corrigan gene determination and therapeutic screening using signature			

Qy 607 TCTAGTGGCTGTATTTCTTACTTATCTTCAATTTTGGACCTGACAGAAATATAGC 666
 Db 613 TCTACTGGCTGTATTTCTTACTTATCTTCAATTTTGGACCTGACAGAAATATAGC 672
 Qy 667 CCATAAATTCACACCTGGAGGAGTGTGTTTGGAGAGGATATGA-TTTTATGAGAAAT 725
 Db 673 ACATGAATTTACCACTGGAGGAGTGTGTTTGGAGAGAGATGATTTTATGAGAGGG 732
 Qy 726 GATATGGCAATGTGCT--AACGATTTTGTATGAAAAGT-TTCCAGACTACTTCTACAG 782
 Db 733 GGTACAGCAAGTGTGATCAGATTTGATATTAAGTACTTTTAAGTACTTCCACAG 792
 Qy 783 TATTTGGTCAATTTTGGAAATGCTTTAGTCTTCACTTTAAATTTATGCTAA 842
 Db 793 TCTTTGGTCAATTTTGAATG-----TCTTCTTCACTTTTAAATGATCTCAATTA 846
 Qy 843 CTTTGTATGAGTTCAATTAATTTTGAATGCTAAATGTAATA 882
 Db 847 CTTTGTATGAGTTGAATTAATTTTGAATGTAATAATA 886

RESULT 7
 AF364070 936 bp mRNA linear ROD 04-MAY-2001
 LOCUS Mus musculus SMPX protein (Smpx) mRNA, complete cds.
 DEFINITION AF364070
 ACCESSION AF364070.1 GI:13940507
 VERSION
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.
 TITLE Identification, mapping, and genomic structure of a novel 1 (bases 1 to 936)
 JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
 MEDLINE 20065879
 PUBMED 10598820
 REFERENCE 2 (bases 1 to 936)
 AUTHORS Patzak, D.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang., Fleischmannstr. 42-44, D-17487 Greifswald, Germany

FEATURES
 source
 1..936
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1..936
 /gene="Smpx"
 139..456
 /gene="Smpx"
 /codon_start=1
 /product="SMPX protein"
 /protein_id="AAK50398.1"
 /db_xref="GI:13940508"
 /translation="MSKQPISNVRAIOANINIMGAFRPGAGPPRRKSTPTEEGA
 PTTSEKKPIPKMKPFPGPVNLSIIONVSEDKFVKHQ"
 906..911
 /gene="Smpx"
 /evidence="experimental"

BASE COUNT 283 a 187 c 203 g 263 t
 ORIGIN
 Query Match 52.0%; Score 461; DB 10; Length 936;
 Best Local Similarity 74.2%; Pred. No. 2.1e-103;
 Matches 689; Conservative 0; Mismatches 180; Indels 60; Gaps 6;

Qy 5 CTCAATACCGGAGAGACAGACTATTTTGGCAATGAAAAGCATGGAAATGAGAT 64
 Db 11 CTGAGAGAGACAGACACTTCCAGCTATTTTCAGCCACATGAAAAGCATGGAAATGAGAT 70

Qy 65 GCGAGCTCAGAGACACCGGGGCGCCCTTCCACTTCCAGAGAC--TTTGTATCTTGC 122
 Db 71 CCCGCTCAGAGACACCGGAGTTCCTTATCTGTAAGCGCTTTTGTGTTTTGC 130
 Qy 123 ATCTGGCTCCTGGGACTTCCCTTAGGCAATGAAATCATTAAGCAGGATTAAGACT 182
 Db 131 ACCTGGCGGCTGGGAGCTGTCTCAGGCAATGAAATCATTAAGCAGGCTTAAGACC 190
 Qy 183 GCATGATATGTGCAAAACGACGATTTCCATGTTAAGCCATTCAGGCAATATCAATA 242
 Db 191 TTGTGAATATGTCCAGACAGCAATTTCCAACTCAGAGCATTCAGGCAATATCAATA 250
 Qy 243 TTCCAAATGGAGGCTTTTCGGCAGAGAGAGTCAACCCCGAGAAAGAAATGATCTC 302
 Db 251 TTCCAAATGGAGGCTTTTCGGCAGAGAGAGTCAACCCCGAGAAAGAAATGATCTC 310
 Qy 303 CTGAAGTGGAGAGGAGTGTTCCTCCACTCGATGAGAGAGAAAGCCAAATTCAGAG 362
 Db 311 CTGAAGTGGAGAGGAGTGTTCCTCCACTCGATGAGAGAGAAAGCCAAATTCAGAG 367
 Qy 363 CGAAGAAATCTTCAGAGCTTGACGATCAATCTATCGAAATTCAGAAATTTAAAGTGAAC 422
 Db 368 TGAAGAAATTTCCAGAGCTGTGTCACTGTCTGATCTGAGATCCAAATGTTAAAGTGAAC 427
 Qy 423 TAAATATGTCCTCCAAAGCTGAACAGTATGAGAAAGAAAGATTTGATGTAAGAAAT 482
 Db 428 TGAATATGTCCTCCAAAGGAGTGAACAGTATGAGAAAGCAAAAGTTCACATTTGATGC 487
 Qy 483 AAGAGGAGAGAGATGAGATTT-----502
 Db 488 TTGAATATGAGAGATGAGATTTGTCGTCGTCGTTTCCAAAGGAGAAAGCAATGGGTT 547
 Qy 503 -----CAATAGCTCAATAAATTTATATATTTATATGATGATGTAAGCTTCGTAAT 556
 Db 548 GAAATTAACAACCTTCGAAACATTTTATCATTTGATGATGATGATCAAAACCTTCGTAAT 607
 Qy 557 GCCGAGACTCTGACAGAAATGAGCTGTGTGATCATTTATATCTCTCTTCTAGTGGC 616
 Db 608 GCCCAAGCTCTGACAGAAATATCTGTTGATCATTTATATTTCTCTTCTTACTTGGT 667
 Qy 617 TGTATTTCTTACTTATCTTCAATTTTGGACCTCAGAAACAAATTTAGGCCATTAATTC 676
 Db 668 TGCATTTCTCACTTACTTACTTATTTTGGCACTTGTAGAGCAATTCAGACAGAAATTT 727
 Qy 677 AACACCTGAGAGGAGTGTGTTTGAAGAGGATATGA-TTTTATGAGAAATGATATGGCAA 735
 Db 728 ACCACCTGGAGAGTGTGCTTTTGAAGAGAGATGATTTTATGAAGAGGGGAGATGGCAA 787
 Qy 736 TGTGCT--CTAAGATTTTGAATGAAAGTTTCCAGAGTACTTCTTCACATTTTGTGTC 793
 Db 788 CGTCAAGACAGATATTTGATGTTAGTACTTTAGTACTTCCACAGGCTTTGTGTC 847
 Qy 794 ATATTTGGAATGCGTTTACTTCTTCACTTTTAAATTTATGACTTAACTTTGTATGAG 853
 Db 848 ATATTTGAATG-----GTTTCTTCACTTTTAAATTTATCTAATTTACTTTTATGAG 901
 Qy 854 TTCAATTAATTAATTTTGAATGTAATAA 882
 Db 902 TTTAAATTAATTAATTTGATGTAATAA 930

RESULT 8
 MMU245772 943 bp mRNA linear ROD 12-APR-2001
 LOCUS Mus musculus mRNA for stretch responsive muscle (X-chromosome)
 DEFINITION MMU245772
 ACCESSION AJ245772
 VERSION AJ245772.1 GI:10178962
 KEYWORDS Srmx gene; stretch responsive muscle (X-chromosome).
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 943)
 AUTHORS Kemp, T.J., Sadusky, T.J., Simon, M., Brown, R., Eastwood, M.,
 Sasseon, D.A. and Coulton, G.R.
 TITLE Identification of a novel stretch-responsive skeletal muscle gene
 (Smpx)
 JOURNAL Genomics 72 (3), 260-271 (2001)
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2 (bases 1 to 943)
 AUTHORS Kemp, T.J.
 TITLE Direct Submission
 JOURNAL Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial
 College School Of Medicine, 6th Building, Exhibition Road, South
 Kensington, London SW7 2AZ, UNITED KINGDOM
 FEATURES
 source
 1..943
 /organism="Mus musculus"
 /strain="C57BL/10"
 /db_xref="taxon:10090"
 /feature_type="skeletal muscle"
 gene
 1..943
 /gene="Smpx"
 1..214
 /gene="Smpx"
 215..472
 /gene="Smpx"
 /codon_start=1
 /product="stretch responsive muscle (X-chromosome)"
 /protein_id="CAC08493.1"
 /db_xref="GI:10178963"
 /translation="MSKOPISNVAIQANINIPMGAPPGAGOPRRKESTPETEGA
 PTSEKKPIPGMKKPPGVNLSIIONVSEKFKVKGHO"
 473..943
 /gene="Smpx"
 535..539
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 585..590
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 658..662
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 706..710
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 740..744
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 779..785
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 909..913
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 923..928
 /gene="Smpx"
 /function="mRNA dectabilising motif"
 943
 /gene="Smpx"
 BASE COUNT 278 a 192 c 210 g 262 t 1 others
 ORIGIN
 Query Match 51.2%; Score 454.4; DB 10; Length 943;
 Best Local Similarity 73.9%; Pred. No. 9.1e-102;
 Matches 684; Conservative 0; Mismatches 181; Indels 61; Gaps 6;
 QY 5 CTCGAATCCGGAGAGACACAGACTATTTACGCCACATGAAAAGCATCGAATTGAGAT 64
 DB 27 CTGGAGAGACACAGAGACTTCACGACTATTTACGCCACATGAAAAGCATCGAATTGAGAT 86
 QY 65 CGCAGCTCAGAGACACCGGCGCCCTTCCACCTTCCAGGAGC--TTTGATTTCTTGC 122
 DB 87 CCCGCTCAGAGACACCGGAGTTCCTTCTATCTGTAAAGCCTTTTGTGTTTGC 146

QY 123 ATTCGCTGCTGGAGCTTCCCTTAGCGATGAACAAATACATTAAGCAGGATTAAGACT 182
 DB 147 ACCTGCGCGCTGGAGCTGTCTTCAGGAGTAACCAATCCAGAGACAGGGCTTAAGCC 206
 QY 183 GCATGAATATGTGAAAACAGCAGCTTCCAAATGTATAGCATTCAGCAATATCAATA 242
 DB 207 TTGTGAATATGTGAAAGAGCAATTTCCAAAGCAGCATTCAGCAATATCAATA 266
 QY 243 TTCAATGGAGCCTTTTCGCGCAGAGCAGATCAACCCCGAGAAAGAAATGTACTC 302
 DB 267 TTCAATGGAGCCTTTTCGCGCAGAGCAGATTCAGCAATTCAGCAATATCAATA 326
 QY 303 CTGAAGTGAAGAGAGGCTTCTCCACCTCGATGAGAAAGAAAGCAATTCAGAG 362
 DB 327 CTGAAGTGAAGAGAGGCTTCTCCACCTCGATGAGAAAGAAAGCAATTCAGAG 383
 QY 363 CGAAGAACTTCCAGAGCCTGAGTCAATCTATCGGAAATCCAGAAATATTAAGTGAAC 422
 DB 384 TGAAGAAATTTCCAGAGCCTGAGTCAATCTGATCGAATCAAAATGTTAAAGTGAAC 443
 QY 423 TAAATATGTCCCAAGCTGAACAGTATGAGAAAGAAAGCAATTCAGAG 482
 DB 444 TGAATTTGCTCCCAAGCTGAACAGTATGAGAAAGCAATTCAGAG 503
 QY 483 AAGAGGCAAGATGATTT----- 502
 DB 504 TTGAATACAGAAATGATTTCTGTCGATTTTCCAGAGAAAGAAAGCAATGAGTT 563
 QY 503 -----CATAGCTCACTAAATTTATATTTATATGATGATGATGAACTTCGAT 556
 DB 564 GAAATTAACAATTCCTGACATTTTATATATGATGATGATGATGATGATGATGATGAT 623
 QY 557 GCTGAGACTCTAGCAAAATGAGCTGTTGATGATGATGATGATGATGATGATGATGAT 616
 DB 624 GCCAAGACTCTAGCAAAATGAGCTGTTGATGATGATGATGATGATGATGATGATGAT 683
 QY 617 TGTATTTCTACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 676
 DB 684 TGTATTTCTACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 743
 QY 677 AACACCTGAGAGGTGTGTTTGAAGAGGATATGATTTATGGA--GAATATATGGA 734
 DB 744 ACAACCTGAGAGGTGTGTTTGAAGAGGATATGATTTATGGAAGGAGGAGGAGGGA 803
 QY 735 ATGTGC--CTAACGATTTGATGAAAAGTTTCCAGACTTCTTCAAGATTTTGTGTC 792
 DB 804 ACGTGCAAGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 863
 QY 793 AATATTTGAAATGCGTTTATGATTTTCACTTCACTTCACTTCACTTCACTTCACTT 852
 DB 864 AATATTTGAAATG-----GTTTCACTTCACTTCACTTCACTTCACTTCACTT 917
 QY 853 GTTCAATTAATATTTGATTAATGT 878
 DB 918 GTTCAATTAATATTTGATTAATGT 943
 RESULT 9
 LOCUS U73508 36503 bp DNA linear PRI 27-APR-1999
 DEFINITION Homo sapiens cosmid clone U1128 from Xp22.1-22.2, complete
 sequence.
 ACCESSION U73508
 VERSION U73508.1 GI:1616808
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 36503)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE The sequence of Homo sapiens cosmid clone U112E8
 JOURNAL Unpublished (1999)
 REFERENCE 3 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-1996)
 REFERENCE 4 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 SUBMITTED BY: WUGSC
 COMMENT Department of Genetics
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 This clone was mapped by Grief, M., Whyte, M. P., Thakker, R. V., and Mazzarella, R. Sequence analysis of 139 Kb in Xp22.1 containing spermine synthase and the 5' region of PEX. Genomics 44:227-231 (1997).

SOURCE INFORMATION:
 This clone is from a chromosome X-specific cosmid library L10XNCC01 'U'. The source of the chromosomes was a human/hamster hybrid, GM07297-F, from Robert Nussbaum at the University of Pennsylvania School of Medicine. Please contact the Lawrence Livermore National Laboratory at http://www-bio.llnl.gov/genome to obtain the clone.
 VECTOR: lawr1st16

FEATURES
 source
 1..36503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp22.1-22.2"
 /clone="U112E8"
 /clone_11b="L10XNCC01-U"
 116..196
 repeat_region
 /rpt_family="L2"
 1980..2046
 repeat_region
 /rpt_family="L2"
 3690..4103
 repeat_region
 /note="match to EST AA211443 (NID:G1810130) zn55b01.b1"
 3711..3929
 repeat_region
 /note="match to EST AA214031 (NID:G1812669) zn58f02.b1"
 complement(3980..4103)
 misc_feature
 /note="match to EST AA211521 (NID:G1810175) zn55b01.r1"

repeat_region 4141..4364
 /rpt_family="L1"
 repeat_region 4366..4730
 /rpt_family="MaLR"
 repeat_region 4732..4953
 /rpt_family="L1"
 repeat_region 5596..5711
 /rpt_family="MaLR"
 repeat_region 6433..6793
 /rpt_family="L1"
 repeat_region 7169..7254
 /rpt_family="L2"
 repeat_region 7874..8264
 /rpt_family="L1"
 repeat_region 8529..8555
 /rpt_family="L2"
 repeat_region 10938..11032
 /rpt_family="MIR"
 repeat_region 11095..11460
 /rpt_family="MaLR"
 repeat_region 11640..11719
 /rpt_family="L1"
 repeat_region 12097..12191
 /rpt_family="L1"
 repeat_region 12453..12595
 /rpt_family="L1"
 repeat_region 13060..13222
 /rpt_family="MIR1_type"
 repeat_region 13286..13385
 /rpt_family="MIR1_type"
 repeat_region 13401..13625
 /rpt_family="MIR1_type"
 repeat_region 14107..14274
 /rpt_family="MIR"
 repeat_region 14638..14751
 /rpt_family="L1"
 repeat_region 14804..15170
 /rpt_family="MaLR"
 repeat_region 15277..15396
 /rpt_family="L2"
 repeat_region 15512..15625
 /rpt_family="L2"
 repeat_region 16941..17238
 /rpt_family="Alu"
 repeat_region 17746..18251
 /rpt_family="MaLR"
 repeat_region 19251..19321
 /rpt_family="MIR1_type"
 repeat_region 19583..19881
 /rpt_family="Alu"
 repeat_region 20137..20192
 /rpt_family="L1"
 repeat_region 20198..20256
 /rpt_family="MIR1_type"
 repeat_region 20456..20562
 /rpt_family="MIR"
 repeat_region 20820..21010
 /rpt_family="MIR1_type"
 repeat_region 23777..23857
 /rpt_family="MIR"
 repeat_region 24315..24471
 /rpt_family="MIR1_type"
 repeat_region 24480..24571
 /rpt_family="MIR"
 repeat_region 24736..24842
 /rpt_family="L2"
 repeat_region 26372..26608
 /rpt_family="MIR"
 repeat_region 27011..27234
 /rpt_family="MIR1_type"
 repeat_region 28149..28286
 /rpt_family="MIR"
 repeat_region 28660..29064

```

repeat_region      /rpf_family="L1"
29526..29750      /rpf_family="L1"
repeat_region      /rpf_family="L1"
30427..30777      /rpf_family="MALR"
repeat_region      /rpf_family="MALR"
30958..30988      /rpf_family="MERL_type"
repeat_region      /rpf_family="MERL_type"
34152..34277      /rpf_family="MIR"
repeat_region      /rpf_family="MIR"
34281..34328      /rpf_family="L2"
m1sc_feature       /note="match to EST AA211443 (NID:g1810130) zn55D01.81"
35221..35386      /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
m1sc_feature       /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
35221..35386      /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
m1sc_feature       /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
35221..35386      /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
repeat_region      /rpf_family="Retroviral"
35277..36501      /rpf_family="Retroviral"
BASE COUNT      10773 a 7354 c 7510 g 10866 t
ORIGIN
Query Match      47.8%; Score 424; DB 9; Length 36503;
Best Local Similarity 100.0%; Pred. No. 3.3e-94;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
464 AGGATTGATGGAAGAAATTAAGAGGAGAGGAGATTCATAGCTCACTAAATTTTA 523
|||||
4103 AGATTGATGGAAGAAATTAAGAGGAGAGGAGATTCATAGCTCACTAAATTTTA 4044
|||||
524 TATATTGATGATGATGGAAGCTCTGATGCTGAGACTCTAGCAGAAATGGCTG 583
|||||
4043 TATATTGATGATGATGGAAGCTCTGATGCTGAGACTCTAGCAGAAATGGCTG 3984
|||||
584 TTTGATGATTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 643
|||||
3983 TTTGATGATTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3924
|||||
644 GGCACCTCAGAGAAACAATTAAGCCCATTAATTAACAACCTGGAGGATGTTGAGGA 703
|||||
3923 GGCACCTCAGAGAAACAATTAAGCCCATTAATTAACAACCTGGAGGATGTTGAGGA 3864
|||||
704 GGGATATGATTTTATGAGAGATGATGAGCAATGCTCAACATTTTGAATGAAGTTT 763
|||||
3863 GGGATATGATTTTATGAGAGATGATGAGCAATGCTCAACATTTTGAATGAAGTTT 3804
|||||
764 CCCAAGTACTCTCTCAAGATTTTGGTCAATTTTGGATGCTTTTACTTCACT 823
|||||
3803 CCCAAGTACTCTCTCAAGATTTTGGTCAATTTTGGATGCTTTTACTTCACT 3744
|||||
824 TTTAAATTTATGCTCACTAACTTTGATGAGTTCAATTAATTTTGAATGAAGTTT 883
|||||
3743 TTTAAATTTATGCTCACTAACTTTGATGAGTTCAATTAATTTTGAATGAAGTTT 3684
|||||
884 GTGA 887
|||||
Db 3683 GTGA 3680

```

```

RESULT 10
AL772392/c 124186 bp DNA linear PRI 18-JUL-2002
LOCUS Human DNA sequence from clone Rpl1-450P7 on chromosome X, complete
DEFINITION
ACCESSION AL772392
VERSION AL772392.4 GI:21912756
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 124186)
AUTHORS Chapman,J.

```

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (18-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone Requests: clonerequests@sanger.ac.uk
On Jul 19, 2002 this sequence version replaced g1:21614755.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
Rpl1-450P7 is from the library RPl-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.choi.org/bacpac/home.htm>
VECTOR: pBACe3.6.

FEATURES

source 1..124186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RPl1-450P7"
/clone_1fb="RPl-11.2"
BASE COUNT 37400 a 24595 c 23947 g 38244 t
ORIGIN

Query Match 47.8%; Score 424; DB 9; Length 124186;
Best Local Similarity 100.0%; Pred. No. 3.5e-94;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

464 AGGATTGATGGAAGAAATTAAGAGGAGAGGAGATTCATAGCTCACTAAATTTTA 523
|||||
100651 AGGATTGATGGAAGAAATTAAGAGGAGAGGAGATTCATAGCTCACTAAATTTTA 100592
|||||
524 TATATTGATGATGATGGAAGCTCTGATGCTGAGACTCTAGCAGAAATGGCTG 583
|||||
100591 TATATTGATGATGATGGAAGCTCTGATGCTGAGACTCTAGCAGAAATGGCTG 100532
|||||
584 TTTGATGATTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 643
|||||
100531 TTTGATGATTTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 100472
|||||
644 GGCACCTCAGAGAAACAATTAAGCCCATTAATTAACAACCTGGAGGATGTTGAGGA 703
|||||
100471 GGCACCTCAGAGAAACAATTAAGCCCATTAATTAACAACCTGGAGGATGTTGAGGA 100412
|||||
704 GGGATATGATTTTATGAGAGATGATGAGCAATGCTCAACATTTTGAATGAAGTTT 763
|||||
100411 GGGATATGATTTTATGAGAGATGATGAGCAATGCTCAACATTTTGAATGAAGTTT 100352
|||||
764 CCCAAGTACTCTCTCAAGATTTTGGTCAATTTTGAATGCTTTTACTTCACT 823
|||||

```

Db 100351 CCGAAGCTCTTCTCAGAGTATTTGGTCAATTTGGAAATGCCCTTTAGTCTTACCT 100292

QY 824 TTTAAATATGCTACTTAACCTTTGTATGAGTTCAAAATTAATTTTGAATAATGTAAT 883

Db 100291 TTTAAATATGCTACTTAACCTTTGTATGAGTTCAAAATTAATTTTGAATAATGTAAT 100232

QY 884 GTGA 887

Db 100231 GTGA 100228

RESULT 11

AY026524 787 bp mRNA linear ROD 28-JUN-2001

LOCUS AY026524

DEFINITION Mus musculus muscle-specific protein CSL (Csl) mRNA, complete cds.

ACCESSION AY026524.1 GI:14575061

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 787)

Palmer, S., Groves, N., Schindeler, A., Yeoh, T., Biben, C., Wang, C.-C., Sparrow, D.B., Barnett, L., Jenkins, N.A., Copeland, N.G., Koentgen, F., Mohun, T. and Harvey, R.P.

Sparrow, D.B., Barnett, L., Jenkins, N.A., Copeland, N.G., Koentgen, F., Mohun, T. and Harvey, R.P.

The small muscle-specific protein Csl modifies cell shape and promotes myocyte fusion in an insulin-like growth factor 1-dependent manner

J. Cell Biol. 153 (5), 985-998, (2001)

21275706

11381084

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

Submitted (01-FEB-2001) Developmental Biology, Victor Chang Cardiac Research Institute, 384 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia

FEATURES

source

1..787

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/chromosome="X"

1..787

/gene="Csl"

206..463

/gene="Csl"

/note="ortholog of Homo sapiens SMPX"

/codon_start=1

/product="muscle-specific protein CSL"

/protein_id="AAK07682.1"

/db_xref="GI:14575062"

/translation="MSKQPISTNRAIQANINIPMGARPRGAGOPRRKKESTPTETSGA
PTTSEKKPLIPMKPGPVNISEIQNVSEIKFVKGSQ"

BASE COUNT 233 a 167 c 185 g 202 t

ORIGIN

Query Match 43.9%; Score 389.8; DB 10; Length 787;

Best Local Similarity 73.5%; Pred. No. 8.5e-86;

Matches 562; Conservative 0; Mismatches 152; Indels 51; Gaps 3;

QY 5 CTCATATCCGGAGAGACAGAGCTATTTGAGCCACATGAAAGCATCGAATGAGAT 64

Db 18 CTGAGAGAGACAGACAGCTCCAGCTATTTGAGCCACATGAAAGCATCGAATGAGAT 77

QY 65 CGAGCTGAGAGACAGCGGCGGCCCTTCACCTTCACAGAGC--TTTGTATTCCTGC 122

Db 78 CCCGCTCAGAGACAGCGGAGGTTCTTATCTGTAAGCGGCTTTTGTGTTTTCG 137

QY 123 ATCTGGCTCCTGGAGCTTCCCTTAGGAGATTAACAATACTAAGAGGATTAAGACT 182

Db 138 ACCTGGCGGCTGGAGCTGCTCTCAGAGATTAACCAATCCAGAGAGGGCTTAAGACC 197

QY 183 GCATGAATATGTCGAAACAGCCAGCTTTCATGTTAGAGCATCAGGCAATATCATATA 242

Db 198 TTGTGAATATGTCGAAACAGCCAGCTTTCATGTTAGAGCATCAGGCAATATCATATA 257

QY 243 TTCAATGAGAGCTTTTGGGCGAGAGCAGTCAACCCCGAGAGAAAGATGTAATC 302

Db 258 TTCCAAATGGAGCTTTTGTGTCGGAGAGCTGGCGAGCTCCAGAGAGAAAGAGTACTC 317

QY 303 CTGAAGTGGAGAGGGTGTCTCCGACCTCGAGTGAAGAGAAAGCAATTCAGAG 362

Db 318 CTGAAGTGGAGAGGGTGTCTCCGACCTCGAGTGAAGAGAAAGCAATTCAGAG 374

QY 363 CGAAGAACTTCCAGAGAGCTGCAATCATGTCGAATTCAGAAATTCAGAAATTCAGAG 422

Db 375 TGAAGAAATTTCCAGAGAGCTGTCATCTGTCAGATCCAAATGTTAAAGTGAAC 434

QY 423 TAAATATGTCCTCCAAAGCTGAACAGTATGAGAAAGAAAGAGATTGATGAGAAAT 482

Db 435 TGAATATGTCCTCCAAAGCTGAACAGTATGAGAAAGAAAGAGATTGATGATGC 494

QY 483 AAGAGGAGAGAGATGATCAATAGC----- 509

Db 495 TTGAATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 554

QY 510 -----TCAATAAATTTTATATATTTGATGATGATGATGATGATGATGATGATG 556

Db 555 GAATTAACAACCTTCGAAACATTTTATATGATGATGATGATGATGATGATGATGATG 614

QY 557 GCCTGAGACTCTAGAGAAATGCGCTGTTGTATGATGATGATGATGATGATGATGATG 616

Db 615 GCCAAGACTCTAGAGAAATGCGCTGTTGTATGATGATGATGATGATGATGATGATG 674

QY 617 TGATTTCTACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTT 676

Db 675 TGCAATTTCTACTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCTTATCT 734

QY 677 AACACCTGAGAGGAGTGTGTTTGGAGAGATGATGATGATGATGATGATGATGATG 721

Db 735 AACACCTGAGAGAGTGTGTTTGGAGAGATGATGATGATGATGATGATGATGATG 779

RESULT 12

AX322783 290 bp DNA linear PAT 07-JAN-2002

LOCUS AX322783

DEFINITION Sequence 27 from Patent WO0192567.

ACCESSION AX322783

VERSION AX322783.1 GI:18093762

KEYWORDS

SOURCE

ORGANISM

unidentified.

unclassified.

unclassified.

REFERENCE

1 Bunk, D., Reuner, B., Beck, J. and Henkel, T.

AUTHORS

TITLE

Novel target genes for diseases of the heart

JOURNAL

Patent: WO 0192567-A 27 06-DEC-2001;

Medigene AG (DE)

FEATURES

source

1..290

/organism="unidentified"

/db_xref="taxon:32644"

BASE COUNT 103 a 50 c 69 g 68 t

ORIGIN

Query Match 32.7%; Score 290; DB 6; Length 290;

Best Local Similarity 100.0%; Pred. No. 3.9e-61;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 299 ACTCTGAAGTGAAGAGGGTGTCTCCGACCTCGAGTGAAGAGAGCAATTCACA 358

Db 1 ACTCTGAAGTGAAGAGGGTGTCTCTCCGACCTCGAGTGAAGAGAGCAATTCACA 60

```

OY 359 GGAGCGAAGAACTTCCAGACCTGCACTCATCTATGGAATCCAGAAATTTAAAGT 418
Db 61 GGAGCGAAGAACTTCCAGACCTGCACTCATCTATGGAATCCAGAAATTTAAAGT 120
OY 419 GAATTAATAATGTGCCCCAAAGCTGACAGATGATGAGAAAGAAAGATGATGAG 478
Db 121 GAATTAATAATGTGCCCCAAAGCTGACAGATGATGAGAAAGAAAGATGATGAG 180
OY 479 AATAAGAGGAGAGATGATGATCAATAGCTCAATAATTTATATTTATGATGATG 538
Db 181 AATAAGAGGAGAGATGATGATCAATAGCTCAATAATTTATATTTATGATGATG 240
OY 539 ATTGTGAACCTCTGAAATGCTGAGACTGTAGCAGAAATGCGCTGTTGT 588
Db 241 ATTGTGAACCTCTGAAATGCTGAGACTGTAGCAGAAATGCGCTGTTGT 290

RESULT 13
AL732396 228031 bp DNA linear HTG 17-AUG-2002
LOCUS Mus musculus chromosome X clone RP23-93M14, *** SEQUENCING IN
DEFINITION PROGRAMS ***, 8 unordered pieces.
ACCESSION AL732396 GI:22415936
VERSION AL732396.6
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Bukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 228031)
AUTHORS Heath, P.
JOURNAL Direct Submission
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22204493.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DM93M14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 99% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Consensus quality: 224691 bases at least Q40
Consensus quality: 226148 bases at least Q30
Consensus quality: 226956 bases at least Q20
Insert size: 227331; sum-of-contigs
Insert size: 194478; 8.5% error; agarose-fp
Quality coverage: 6.42x in Q20 bases; sum-of-contigs Quality
coverage: 8.63x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
5545 5644: gap of 100 bp
5645 11121: contig of 5477 bp in length
11122 11221: gap of 100 bp
11222 23110: contig of 11889 bp in length
23111 23210: gap of 100 bp
23211 42146: contig of 18936 bp in length
42147 42246: gap of 100 bp
42247 74925: contig of 32679 bp in length

```

```

* 74926 75025: gap of 100 bp
* 75026 153828: contig of 78803 bp in length
* 153829 153928: gap of 100 bp
* 153929 172635: contig of 18707 bp in length
* 172636 172735: gap of 100 bp
* 172736 228031: contig of 55296 bp in length.
Location/Qualifiers
1. 228031
source /organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-93M14"
/clone_11b="RPC1-23"
1. 5544
/misc_feature /note="assembly fragment:01721
fragment_chain:1"
misc_feature /note="assembly fragment:02676
fragment_chain:1"
11222. 23110
/misc_feature /note="assembly fragment:00239
fragment_chain:1"
23211. 42146
/misc_feature /note="assembly fragment:01463
fragment_chain:1"
42247. 74925
/misc_feature /note="assembly fragment:02519
fragment_chain:1"
75026. 153828
/misc_feature /note="assembly fragment:00456
fragment_chain:1"
153929. 172635
/misc_feature /note="assembly fragment:03454
fragment_chain:2"
172736. 228031
/misc_feature /note="assembly fragment:01243
fragment_chain:2"
BASE COUNT 71213 a 42522 c 42780 g 70814 t 702 others
Query Match 25.0%; Score 221.4; DB 2; Length 228031;
Best Local Similarity 78.3%; Pred. No. 4.4e-44;
Matches 306; Conservative 0; Mismatches 76; Indels 9; Gaps 3;
OY 500 ATTCAATAGCTCACTAAATTTATATATTTGATGATGATGGAACCTCGAATGCC 559
Db 187401 ATAAACAACCTTCGAACATTTATATCATTTGATGATGATGATCAACCACTCGAATGCC 187460
OY 560 TGAGACTCTAGCAGAAATGCGCTGTTGTACATTATATCTCTCTAGTTGCGTGT 619
Db 187461 CAGACTCTAGCAGAAATGCGCTGTTGTACATTATATCTCTCTAGTTGCGTGT 187520
OY 620 ATTCTTACTTTATCTTCATTTTGGCAGCTCACAGAACAAATTAGCCATAATTGCAAC 679
Db 187521 ATTCTTACTTTAGCTACATTTTGGCAGCTTGTAGAGCAATGACACGAAATTTACA 187580
OY 680 ACCGTGAGGGTGTGCTTTTGAAGAGGATATGA-TTTTATGAGAAATGATATGGCAATGT 738
Db 187581 ACCGTGAGAAATGTGCTTTTGAAGAGATGTGATTTTATGAAAGGGGGGATGCAACGT 187640
OY 739 GC--CTAACATTTTGTGAAAGATTCCCAAGCTACTTCTCAGATTTTGGTCATA 796
Db 187641 GCAAGCAGATTTTGTATGTTAGTACTTTAAGTTACTTCCACGGGCTTTGGTCATA 187700
OY 797 TTGTGAATGCGTTTAACTTTCACCTTTTAAATATATGATGATGATGATGATG 856
Db 187701 TTGTGAATG-----GTTTCTCACCTTTTAAATATATCAATTAATCTTTATGAGTTC 187754
OY 857 AATAATAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 887
Db 187755 AATAATAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 187785

```

RESULT 14
AC127606
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC127606 186670 bp DNA linear HTG 18-JUL-2002
Rattus norvegicus clone CH230-20D15, *** SEQUENCING IN PROGRESS
***, 60 unordered pieces.
AC127606
AC127606.1 GI:21902814
HTG, HTGS PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 186670)
Wuzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrook,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,B., Kelly,S., Khan,U., King,L., Korah,U., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokhekwo,S., Ogum,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Dickens,R., Prime,E., Pu,L.L., Quiles,M., Ran,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Schierer,S., Scott,G., Shen,H., Shooshitari,N., Slison,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,I.,
Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Uman,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,D., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL

Unpublished
2 (bases 1 to 186670)
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GXOD
Center clone name: CH230-20D15
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads

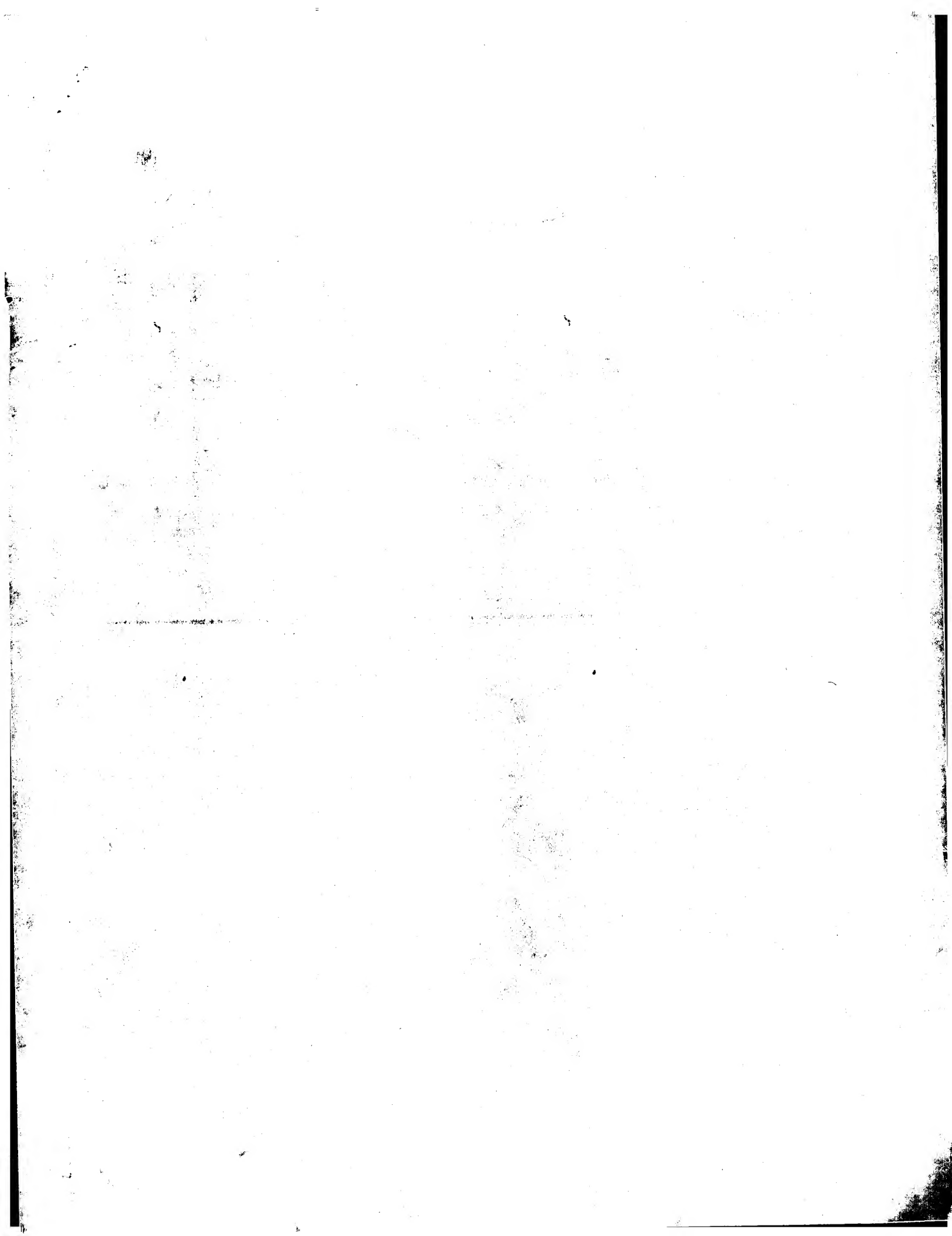
Assembly program: Phrap; version 0.990329
Consensus quality: 142614 bases at least Q40
Consensus quality: 149053 bases at least Q30
Consensus quality: 154556 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1104: contig of 1104 bp in length
* 1105 1204: gap of unknown length
* 1205 2491: contig of 1287 bp in length
* 2492 2591: gap of unknown length
* 2592 3744: contig of 1153 bp in length
* 3745 3844: gap of unknown length
* 3845 5424: contig of 1580 bp in length
* 5425 5524: gap of unknown length
* 5525 6848: contig of 1324 bp in length
* 6849 6949: gap of unknown length
* 6949 8220: contig of 1272 bp in length
* 8221 8320: gap of unknown length
* 8321 9618: contig of 1298 bp in length
* 9619 9718: gap of unknown length
* 9719 10979: contig of 1261 bp in length
* 10980 11079: gap of unknown length
* 11080 12189: contig of 1110 bp in length
* 12190 12889: gap of unknown length
* 12890 13796: contig of 1507 bp in length
* 13797 13896: gap of unknown length
* 13897 15402: contig of 1506 bp in length
* 15403 15502: gap of unknown length
* 15503 16696: contig of 1194 bp in length
* 16697 16796: gap of unknown length
* 16797 18391: contig of 1595 bp in length
* 18392 18491: gap of unknown length
* 18492 20949: contig of 2358 bp in length
* 20950 23047: gap of unknown length
* 23048 23147: contig of 2098 bp in length
* 23148 25634: gap of unknown length
* 25635 25734: gap of unknown length
* 25735 26889: contig of 1155 bp in length
* 26890 26990: gap of unknown length
* 26991 29235: contig of 2246 bp in length
* 29236 29335: gap of unknown length
* 29336 30805: contig of 1470 bp in length
* 30806 30905: gap of unknown length
* 30906 32225: contig of 1320 bp in length
* 32226 32325: gap of unknown length
* 32326 34307: contig of 1982 bp in length
* 34308 34407: gap of unknown length
* 34408 35942: contig of 1535 bp in length
* 35943 36042: gap of unknown length
* 36043 38117: contig of 2075 bp in length
* 38118 38217: gap of unknown length
* 38218 40202: contig of 1985 bp in length
* 40203 40302: gap of unknown length
* 40303 42524: contig of 2222 bp in length
* 42525 44620: gap of unknown length
* 44621 44620: contig of 1896 bp in length
* 44621 44620: gap of unknown length
* 44621 47711: contig of 3091 bp in length
* 47712 47811: gap of unknown length
* 47812 50162: contig of 2351 bp in length
* 50163 50262: gap of unknown length
* 50263 52243: contig of 1981 bp in length
* 52244 52343: gap of unknown length


```
OY 560 TGAGACTCTAC -AGAAATGCGCTGTTGTACATTATATCTCTTCCCTTACGTGCTG 618
Db 52951 CAGACTCTAGCAAAAATATCCGTGTTGTACATTATATTTCTTCTTACTTGCTG 52892
OY 619 TATTTCTTACTTATCTTCAATTTTGGCAGCTCAGACAAATTAACCATTAATTCA 678
Db 52891 CATTTCACCTTAGCTACATTTTGGCAGCTTGCAGAGCAAAATGAGCACATGAATTTAC 52832
OY 679 CACCTGAGGGGTGTGTTTGTAGAGAGGATATGA -TTTATGGAGATGATATGCAATG 737
Db 52831 CACCTGCGAGTGTGTTTGTAGAGAGATGATTTTATGAGGGGGGTACAGCAATG 52772
OY 738 TGCCT--AACGATTTGATGAAAGT-TTCCCAAGCTACTTCTACAGTATTTGGTCA 794
Db 52771 TGCATGAGATGATTTGATATTAAGTACTTTAACTTACTCCACAGTCTTTGGTCA 52712
OY 795 TATTTGAAATGCGTTTATGTTCTTCACTTTTAATTATGTCATAACTTTGTATGAGT 854
Db 52711 TATTTGAAATG-----TCTTCTTCACTTTTAATGATCTCAATTAACCTTTTATGAGT 52658
OY 855 TCAATTAATATTTTGACTAATGTAAATGTGA 887
Db 52657 TGAATTAATATTTTGAATGAATGTAAATGTATTA 52625
```

Search completed: April 15, 2003, 20:49:22
Job time : 2441.65 secs



GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:33:56 ; Search time 16.6774 Seconds
(without alignments)
155.253 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459

Sequence: 1 MNMSKQPVSNVRAIQANINI.....NLSEIQIKSELKVPKAEQ 88

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA: *
2: /cgn2_6/ptodata/1/1aa/5a COMB pep: *
3: /cgn2_6/ptodata/1/1aa/5b COMB pep: *
4: /cgn2_6/ptodata/1/1aa/6a COMB pep: *
5: /cgn2_6/ptodata/1/1aa/6b COMB pep: *
6: /cgn2_6/ptodata/1/1aa/6c COMB pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.5	16.7	1274	4 US-09-095-443-2	Sequence 2, Appl1
2	73.5	16.0	659	4 US-08-426-509A-8	Sequence 8, Appl1
3	73.5	16.0	659	PCT-US95-05008-8	Sequence 8, Appl1
4	72	15.7	392	3 US-08-301-162-2	Sequence 2, Appl1
5	72	15.7	392	4 US-09-461-240-2	Sequence 2, Appl1
6	72	15.7	392	4 US-08-968-927-2	Sequence 2, Appl1
7	72	15.7	428	3 US-08-301-162-18	Sequence 18, Appl1
8	72	15.7	428	4 US-09-461-240-18	Sequence 18, Appl1
9	72	15.7	428	4 US-09-968-927-18	Sequence 18, Appl1
10	70.5	15.4	257	2 US-08-414-926A-7	Sequence 7, Appl1
11	70.5	15.4	257	2 US-08-926-922-7	Sequence 7, Appl1
12	70.5	15.4	257	2 US-08-253-682-7	Sequence 7, Appl1
13	70.5	15.4	257	4 US-08-527-657-7	Sequence 7, Appl1
14	70	15.3	667	4 US-09-303-064-55	Sequence 55, Appl1
15	69	15.0	872	1 US-08-491-357-3	Sequence 3, Appl1
16	69	15.0	872	1 US-08-968-633-3	Sequence 3, Appl1
17	69	15.0	872	3 US-08-968-633-3	Sequence 3, Appl1
18	69	15.0	872	3 US-09-196-466-3	Sequence 3, Appl1
19	69	15.0	872	PCT-US96-10823-3	Sequence 3, Appl1
20	68	14.8	243	1 US-07-960-112B-2	Sequence 2, Appl1
21	68	14.8	243	1 US-08-301-316B-2	Sequence 2, Appl1
22	68	14.8	243	1 US-08-473-399B-2	Sequence 2, Appl1
23	68	14.8	243	3 US-08-853-831-2	Sequence 2, Appl1
24	68	14.8	243	5 PCT-US93-09774-2	Sequence 2, Appl1
25	67.5	14.7	442	1 US-08-391-615-6	Sequence 6, Appl1
26	67.5	14.7	2441	1 US-08-194-468-2	Sequence 2, Appl1
27	67.5	14.7	2441	3 US-08-961-739-2	Sequence 2, Appl1

28	67.5	14.7	2441	4 US-09-514-247A-8	Sequence 8, Appl1
29	66	14.4	708	4 US-09-413-814-18	Sequence 18, Appl1
30	65.5	14.3	66	3 US-09-020-116-1	Sequence 1, Appl1
31	65.5	14.3	66	4 US-09-608-902-1	Sequence 1, Appl1
32	65.5	14.3	512	4 US-08-426-509A-16	Sequence 16, Appl1
33	65.5	14.3	512	5 PCT-US95-05008-16	Sequence 16, Appl1
34	65.5	14.3	659	1 US-08-391-615-2	Sequence 2, Appl1
35	65	14.2	356	4 US-08-887-534A-72	Sequence 72, Appl1
36	65	14.2	999	2 US-08-770-301A-1	Sequence 1, Appl1
37	65	14.2	999	3 US-09-175-581-1	Sequence 1, Appl1
38	64.5	14.1	212	4 US-08-973-462-25	Sequence 25, Appl1
39	64.5	14.1	468	4 US-09-537-357-53	Sequence 53, Appl1
40	64.5	14.1	630	4 US-08-973-462-9	Sequence 9, Appl1
41	64	13.9	999	2 US-08-770-301A-3	Sequence 3, Appl1
42	64	13.9	999	3 US-09-175-581-3	Sequence 3, Appl1
43	63.5	13.8	264	3 US-08-894-731-4	Sequence 4, Appl1
44	63	13.7	436	1 US-08-080-255-8	Sequence 8, Appl1
45	63	13.7	436	3 US-08-465-713-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-09-095-443-2
; Sequence 2, Application US/09095443
; Patent No. 6342593
; GENERAL INFORMATION:
; APPLICANT: Plowman, Gregory
; APPLICANT: Peles, Elor
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF ALP RELATED DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; MEDIUM TYPE: storage
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,443
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/049,477
; FILING DATE: June 12, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEPHONE: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1274 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-095-443-2
Query Match 16.7%; Score 76.5; DB 4; Length 1274;
Best Local Similarity 27.0%; Pred. No. 3.2;

Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;

QY 7 PVSIVRAIQANINI-PMG-----AFRPGAGQPPRRKCTPEVEBGVPTSDDEKKPIPG 59

Db 1132 PISIOATIAIATLSIRPGLSPVSLPGPABPP-----GLPPALPESTPIPS 1180

QY 60 AKLPGAVNLSEIQNIKSELKYVPKAEQ 88

Db 1181 SSPP-----LSSPLPEADQPKK 1198

RESULT 2

US-08-426-509A-8
Sequence 8, Application US/08426509A
Patent No. 6326469

GENERAL INFORMATION:

APPLICANT: Ullrich, Axel

APPLICANT: Gishizaky, Mikhail

APPLICANT: Sures, Itman G.

TITLE OF INVENTION: NOVEL MEGAKARYOCYTIC PROTEIN

TITLE OF INVENTION: TYROSINE KINASES

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York,

STATE: NY

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/426,509A

FILING DATE: 21-APR-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/232,545

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7683-0074-999

TELEPHONE: 212-790-9090

TELEFAX: 212-869-9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 659 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: No. 6326469e

US-08-426-509A-8

Query Match

Best Local Similarity 32.4%; Pred. No. 3.1;

Matches 22; Conservative 13; Mismatches 14; Indels 19; Gaps 5;

QY 23 GARPRGAGQPPRRKCTPEVEBGVPTSDDE--KKPIGAKLPGPA--VNLSEIQNI 76

Db 173 GSLKPGSS--HRTKTP-----LPTPEBDQILKKPLP-----PEPAAPVSTSEIKV 219

QY 77 KSELKYVP 84

Db 220 VALDYMP 227

RESULT 3

PCT-US95-05008-8

Sequence 8, Application PC/TUS9505008

GENERAL INFORMATION:

APPLICANT: Sugen, Inc.

APPLICANT: 515 Galveston Drive

APPLICANT: Redwood City, California 94063-4720

APPLICANT: United States of America

APPLICANT: Wissenschaften E.V.

APPLICANT: Hofgarten Str. 2

APPLICANT: München 80539

APPLICANT: Germany

TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine

TITLE OF INVENTION: Kinases

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/05008

FILING DATE: 24-APR-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/232,545

FILING DATE: 22-APR-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7683-074

TELEPHONE: (212)790-9090

TELEFAX: (212)869-9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 659 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

PCT-US95-05008-8

Query Match

Best Local Similarity 32.4%; Pred. No. 3.1;

Matches 22; Conservative 13; Mismatches 14; Indels 19; Gaps 5;

QY 23 GARPRGAGQPPRRKCTPEVEBGVPTSDDE--KKPIGAKLPGPA--VNLSEIQNI 76

Db 173 GSLKPGSS--HRTKTP-----LPTPEBDQILKKPLP-----PEPAAPVSTSEIKV 219

QY 77 KSELKYVP 84

Db 220 VALDYMP 227

RESULT 4

US-08-301-162-2

Sequence 2, Application US/08301162

Patent No. 6022546

GENERAL INFORMATION:

APPLICANT: Knapp, Stefan

APPLICANT: Ziegelmaler, Robert

APPLICANT: Kupper, Hans

TITLE OF INVENTION: Tokoplasma Gondi Antigen. The

Preparation Thereof and the Use Thereof

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/167,128
FILING DATE:
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2
FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleisher, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02481.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-301-162-2

Query Match 15.7%; Score 72; DB 3; Length 392;
Best Local Similarity 24.1%; Pred. No. 2.4;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSKQPVSNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPPTSDEKKPIPGAK 61
DB 164 NAGELPPPNVQEL-----PPPNVQELPPPTGEQLPPPTGEQLPP-PTGQ 206
QY 62 KLPGPAVNLSEIQNIKSELYKVPKAEQ 88
DB 207 ELPPP-----TEQELAPSTEQ 222

RESULT 5:
US-09-461-240-2
Sequence 2, Application US/09461240
Patent No. 6326008
GENERAL INFORMATION:
APPLICANT: Knapp, Stefan
Kupper, Hans
Ziegelmaier, Robert
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
Preparation Thereof and the Use Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/461,240
FILING DATE: 16-DEC-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE: <Unknown>
APPLICATION NUMBER: US/08/167,128
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2
FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleisher, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02481.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-461-240-2

Query Match 15.7%; Score 72; DB 4; Length 392;
Best Local Similarity 24.1%; Pred. No. 2.4;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSKQPVSNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPPTSDEKKPIPGAK 61
DB 164 NAGELPPPNVQEL-----PPPNVQELPPPTGEQLPPPTGEQLPP-PTGQ 206
QY 62 KLPGPAVNLSEIQNIKSELYKVPKAEQ 88
DB 207 ELPPP-----TEQELAPSTEQ 222

RESULT 6:
US-09-968-927-2
Sequence 2, Application US/09968927
Patent No. 6419925
GENERAL INFORMATION:
APPLICANT: Knapp, Stefan
Kupper, Hans
Ziegelmaier, Robert
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
Preparation Thereof and the Use Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/968,927
FILING DATE: 03-Oct-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE: <Unknown>
APPLICATION NUMBER: US/08/167,128
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2
FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleishner, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02481.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-968-927-2

Query Match 15.7%; Score 72; DB 4; Length 392;
Best Local Similarity 24.1%; Pred. No. 2.4;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSKQPSNVRAIOANINIMGAFRPGAGPPRRKCTPEVEGVPPTSDEKKPIPGAK 61
DB 164 NAQELPPNVQEL-----PPNVQELPPPTQELPPTQELPP-PTQ 206
QY 62 KLPGPAVNLSIQIKSELKTVPAEQ 88
DB 207 ELPPP-----TEQELAPSTEQ 222

RESULT 7
US-08-301-162-18
Sequence 18, Application US/08301162
Patent No. 6022546
GENERAL INFORMATION:
APPLICANT: Knapp, Stefan
APPLICANT: Ziegelmaier, Robert
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
TITLE OF INVENTION: Preparation Thereof and the Use Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pinnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/167,128
FILING DATE:
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2

FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleishner, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02481.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-301-162-18

Query Match 15.7%; Score 72; DB 3; Length 428;
Best Local Similarity 24.1%; Pred. No. 2.6;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSKQPSNVRAIOANINIMGAFRPGAGPPRRKCTPEVEGVPPTSDEKKPIPGAK 61
DB 156 NAQELPPNVQEL-----PPNVQELPPPTQELPPTQELPP-PTQ 198
QY 62 KLPGPAVNLSIQIKSELKTVPAEQ 88
DB 199 ELPPP-----TEQELAPSTEQ 214

RESULT 8
US-09-461-240-18
Sequence 18, Application US/09461240
Patent No. 632608
GENERAL INFORMATION:
APPLICANT: Knapp, Stefan
APPLICANT: Ziegelmaier, Robert
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
TITLE OF INVENTION: Preparation Thereof and the Use Thereof
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pinnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/461,240
FILING DATE: 16-DEC-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE: <Unknown>
APPLICATION NUMBER: US/08/167,128
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2
FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleishner, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02481.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-461-240-18

Query Match 15.7%; Score 72; DB 4; Length 428;
Best Local Similarity 24.1%; Pred. No. 2.6;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSQPVSNVRAIQANINIPGAFRPGAGPPRRKCTPEVEEGVPTSDKKPIPGAK 61
DB 156 NAQELPPPNVQEL-----PPPNVQELPPPTQELPPPTQELPP-PTQ 198
QY 62 KLGPANVLSIQNIKSLKXVPRAG 88
DB 199 ELPPP-----TEQELAPSTEQ 214

RESULT 9
US-09-968-927-18
Sequence 18, Application US/09968927
Patent No. 6419925
GENERAL INFORMATION:
APPLICANT: Knapp, Stefan
Ziegelmaier, Robert
Kupper, Hans
TITLE OF INVENTION: Toxoplasma Gondii Antigens, The
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Funnegan, Henderson, Farabow, Garrett &
Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/968,927
FILING DATE: 03-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/301,162
FILING DATE: <Unknown>
APPLICATION NUMBER: US/08/167,128
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/623,086
FILING DATE: 06-DEC-1990
APPLICATION NUMBER: DE P3940598.2
FILING DATE: 08-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Fleisher, Raz E.
REGISTRATION NUMBER: 34,331
REFERENCE/DOCKET NUMBER: 02461.1005-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 428 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 18:

US-09-968-927-18

Query Match 15.7%; Score 72; DB 4; Length 428;
Best Local Similarity 24.1%; Pred. No. 2.6;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3;

QY 2 NMSQPVSNVRAIQANINIPGAFRPGAGPPRRKCTPEVEEGVPTSDKKPIPGAK 61
DB 156 NAQELPPPNVQEL-----PPPNVQELPPPTQELPPPTQELPP-PTQ 198
QY 62 KLGPANVLSIQNIKSLKXVPRAG 88
DB 199 ELPPP-----TEQELAPSTEQ 214

RESULT 10
US-08-414-926A-7
Sequence 7, Application US/08414926A
Patent No. 5721354
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-2155
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414,926A
FILING DATE: March 31, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-011/COUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: col.01
FEATURE:
NAME/KEY: Protein
LOCATION: 1..257
OTHER INFORMATION: /label= UL133
US-08-414-926A-7

Query Match 15.4%; Score 70.5; DB 1; Length 257;
Best Local Similarity 36.8%; Pred. No. 2.1;
Matches 25; Conservative 4; Mismatches 26; Indels 13; Gaps 4;

QY 26 RQAGQPPRRKCTPEVEEGVPTSDKKPIPGAKXKLGPANVLSIQNIK 77
DB 126 RGHGDDPDPRVICIYESPPVAPTVSVPPESHSQVPIP-----PQPPATSPKPKX 180
QY 78 SLKXVPR 85
DB 181 GRAKDKPK 188

RESULT 11
US-08-926-922-7
Sequence 7, Application US/08926922
Patent No. 5925751
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Luann Cseert Attorney at Law
STREET: 750 Arimo Avenue
CITY: Oakland
STATE: CA
COUNTRY: USA
ZIP: 94610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/926,922
FILING DATE: September 10, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR 11A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-834-1448
TELEFAX: 510-834-7810
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: col.01
FEATURE:
NAME/KEY: Protein
LOCATION: 1..257
OTHER INFORMATION: /label=UL133
US-08-926-922-7
Query Match 15.4%; Score 70.5; DB 2; Length 257;
Best Local Similarity 36.8%; Pred. No. 2.1;
Matches 25; Conservative 4; Mismatches 26; Indels 13; Gaps 4;
QY 26 RFGAGQPPRRKEC---TPEVEE---GVPTSDEKKP-IPGAKKLPAPVNLSEIQNIK 77
DB 126 RFGHGDDEDRVCEIVESPVSAPTWSVPSPSESHQVIR-----PQPPAPTSEPKPKK 180
QY 78 SELKYVPK 85
DB 181 GRADKPK 188
RESULT 12
US-09-253-682-7
Sequence 7, Application US/09253682
Patent No. 6040170
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Luann Cseert Attorney at Law
STREET: 750 Arimo Avenue
CITY: Oakland

STATE: CA
COUNTRY: USA
ZIP: 94610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/253,682
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/926,922
FILING DATE: September 10, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR 11A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-834-1448
TELEFAX: 510-834-7810
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 257 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: col.01
FEATURE:
NAME/KEY: Protein
LOCATION: 1..257
OTHER INFORMATION: /label=UL133
US-09-253-682-7
Query Match 15.4%; Score 70.5; DB 3; Length 257;
Best Local Similarity 36.8%; Pred. No. 2.1;
Matches 25; Conservative 4; Mismatches 26; Indels 13; Gaps 4;
QY 26 RFGAGQPPRRKEC---TPEVEE---GVPTSDEKKP-IPGAKKLPAPVNLSEIQNIK 77
DB 126 RFGHGDDEDRVCEIVESPVSAPTWSVPSPSESHQVIR-----PQPPAPTSEPKPKK 180
QY 78 SELKYVPK 85
DB 181 GRADKPK 188
RESULT 13
US-09-527-657-7
Sequence 7, Application US/09527657
Patent No. 6291236
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Luann Cseert Attorney at Law
STREET: 750 Arimo Avenue
CITY: Oakland
STATE: CA
COUNTRY: USA
ZIP: 94610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/527,657
FILING DATE: 17-Mar-2000


```

Query Match      15.4%; Score 70.5; DB 4; Length 257;
Best Local Similarity 36.8%; Pred. No. 2.1;
Matches 25; Conservative 4; Mismatches 26; Indels 13; Gaps 4;

QY      26  RPPGAGGPPRRKEC---TPREVE--GVPPISDEKKP-IGAKKLPQPAVNLSEIQNIK 77
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      126  RFGHGDDPDDVIVICEIVSPVSAPTVSVPPPEESHQVIE-----PQPAPTSEPKPK 180
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      78  SELKTVPK 85
      |||  |||
Db      181  GRAKDKPK 188

```

Query Match 15.3%; Score 70; DB 4; Length 667;
Best Local Similarity 24.1%; Pred. No. 7.7;
Matches 21; Conservative 13; Mismatches 25; Indels 28; Gaps 3

```

RESULT 15
US-09-086-503--55
Sequence 55, Application US/09086503A
Patent No. 6329157
GENERAL INFORMATION:
APPLICANT: MAINE, Gregory T.
APPLICANT: HUNT, Jeffery C.
APPLICANT: BROJANAC, Susan
APPLICANT: JYH-TSING SHEU, Michael
APPLICANT: CHOVAN, Linda E.
APPLICANT: TYNER, Joan D.
TITLE OF INVENTION: ANTIGEN COCKTAILS AND USES THEREOF
FILE REFERENCE: 6361.US. 01
CURRENT APPLICATION NUMBER: US/09/086,503A
CURRENT FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 55
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 55
LENGTH: 667
TYPE: PRF
ORGANISM: Toxoplasma gondii
US-09-086-503--55

```

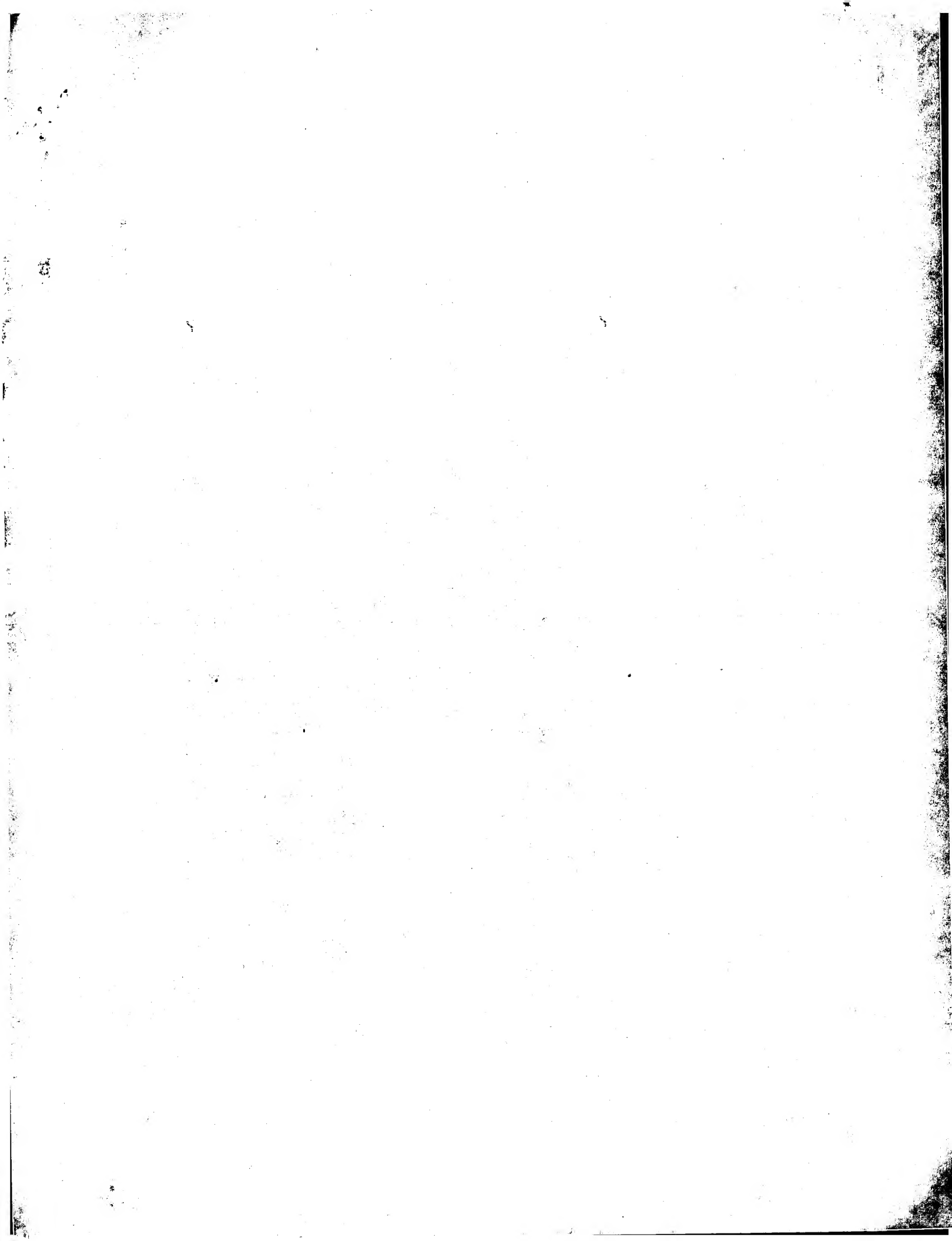
```

Query Match 15.3% Score 70; DB 4; Length 667;
Best Local Similarity 24.1%, Prod. No. 7.7, Indels 28, Gaps
Matches 21; Conservative 11; Mismatches 25;

QY      2  NMSKQSVYVRLIQLNINIPGACRPGAGQPPRRKCTPEVEEGVPTSDKKPIGAK 61
Db      303  NQDELPPPVQGL-----PPNVQELPPPTQELPPPTQELPP-PTQ 345
          |||||
QY      62  KLRGPAVNLSEIQNIKSELKTVPAEQ 88
Db      346  ELPPP-----TEQELPPSTEQ 361
          |||||

```

Search completed: April 9, 2003, 12:40:31
Job time : 18.6774 secs



GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:38:46 ; Search time 19.5161 Seconds

(without alignments)
275.668 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459
Sequence: 1 MNMSKQPVSNVRAIQANINI.....NLSIQNIKSELYKVEKAO 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/us08_NEW_PUB pep: *
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep: *
3: /cgn2_6/ptodata/2/pubpaa/us06_NEW_PUB pep: *
4: /cgn2_6/ptodata/2/pubpaa/us06_PUBCOMB pep: *
5: /cgn2_6/ptodata/2/pubpaa/us07_NEW_PUB pep: *
6: /cgn2_6/ptodata/2/pubpaa/us07_PUBCOMB pep: *
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep: *
8: /cgn2_6/ptodata/2/pubpaa/us08_PUBCOMB pep: *
9: /cgn2_6/ptodata/2/pubpaa/us09_NEW_PUB pep: *
10: /cgn2_6/ptodata/2/pubpaa/us09_PUBCOMB pep: *
11: /cgn2_6/ptodata/2/pubpaa/us10_NEW_PUB pep: *
12: /cgn2_6/ptodata/2/pubpaa/us10_PUBCOMB pep: *
13: /cgn2_6/ptodata/2/pubpaa/us60_NEW_PUB pep: *
14: /cgn2_6/ptodata/2/pubpaa/us60_PUBCOMB pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	18.3	523	10	US-09-799-777-76
2	76.5	16.7	603	10	US-09-906-779-4
3	76.5	16.7	1274	9	US-10-020-215-2
4	73.5	16.0	659	9	US-09-977-260-8
5	73.5	16.0	659	9	US-10-045-202-2
6	73.5	16.0	659	9	US-09-977-261-8
7	73.5	16.0	659	10	US-09-977-269-8
8	70	15.3	667	10	US-09-896-852-55
9	68.5	14.9	978	10	US-09-815-242-5456
10	68.5	14.9	1001	10	US-09-815-242-12686
11	68	14.8	196	10	US-09-858-664A-16
12	68	14.8	247	9	US-09-981-353-95
13	68	14.8	280	10	US-09-925-300-1411
14	67.5	14.7	2441	12	US-10-109-886-8
15	67	14.6	180	10	US-09-864-761-45128
16	67	14.6	437	10	US-09-864-761-36083
17	66.5	14.5	238	10	US-09-864-761-34591
18	66.5	14.5	2005	10	US-09-735-367B-3
19	66.5	14.5	2063	10	US-09-735-367B-2

20	65.5	14.3	108	10	US-09-864-761-35589	Sequence 35589, A
21	65.5	14.3	512	9	US-09-977-260-16	Sequence 16, Appl
22	65.5	14.3	512	9	US-09-977-261-16	Sequence 16, Appl
23	65.5	14.3	512	10	US-09-977-269-16	Sequence 16, Appl
24	65.5	14.3	559	9	US-10-045-202-4	Sequence 4, Appl
25	65	14.2	363	1	US-08-781-986A-5237	Sequence 5237, Ap
26	65	14.2	546	10	US-09-901-884-5	Sequence 5, Appl
27	65	14.2	616	9	US-10-243-735-4	Sequence 4, Appl
28	65	14.2	774	9	US-10-162-706-5	Sequence 5, Appl
29	65	14.2	1530	9	US-10-118-513A-6	Sequence 6, Appl
30	65	14.2	2266	9	US-10-118-513A-14	Sequence 14, Appl
31	64.5	14.1	212	10	US-09-742-096-25	Sequence 12, Appl
32	64.5	14.1	212	10	US-09-925-300-1577	Sequence 25, Appl
33	64.5	14.1	468	10	US-09-884-260A-53	Sequence 1577, Ap
34	64.5	14.1	559	1	US-08-781-986A-5251	Sequence 53, Appl
35	64.5	14.1	630	9	US-09-742-096-5	Sequence 5, Appl
36	63.5	13.8	212	10	US-09-733-507-12	Sequence 1, Appl
37	63.5	13.8	403	10	US-09-731-872-249	Sequence 12, Appl
38	63.5	13.8	466	10	US-09-925-301-1355	Sequence 249, Ap
39	63.5	13.8	930	9	US-10-113-794A-1	Sequence 1355, Ap
40	63	13.7	360	9	US-10-149-819-6	Sequence 1, Appl
41	62.5	13.6	1907	9	US-09-832-292-39	Sequence 6, Appl
42	62	13.5	546	10	US-09-839-743-3	Sequence 39, Appl
43	62	13.5	603	9	US-09-764-868-705	Sequence 3, Appl
44	62	13.5	771	9	US-09-982-107-8	Sequence 705, App
45	62	13.5	808	9	US-10-108-605-127	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-799-777-76
; Sequence 76, Application US/09799777
; Patent No. US20020091244A1
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; Hillman, Jennifer L.
; Guegler, Karl J.
; Baugh, Marian
; Sacher, Susan
; Shah, Puryi
; TITLE OF INVENTION: HUMAN SIGNAL PEPTIDE-CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 154
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,777
; FILING DATE: 06-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/002,485
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BILLINGS, LUCY J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0459 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 845-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 523 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: INDMOT05
; CLONE: 3125156
; SEQUENCE DESCRIPTION: SEQ ID NO: 76 :
US-09-799-777-76

```

```

Query Match          18.3%; Score 84; DB 10; Length 523;
Best Local Similarity 32.8%; Pred. No. 0.52;
Matches 22; Conservative 10; Mismatches 17; Indels 18; Gaps 3;

```

```

QY 7 PVSIVRAIQANINI-PMG-----AFRPGAGQPPRRKCTPEVEBGVPTSDDEKKPIPG 59
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 435 PISIQATIAKLSIRPFGLESFVASLPGPAEP-----GLPPASLPSTPIPS 483
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 60 AKKLPGP 66
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 484 SSQTPFP 490

```

RESULT 2

```

US-09-906-779-4
; Sequence 4; Application US/09906779
; Patent No. US2002006484A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: Human Protein Tyrosine Phosphatase Polynucleotides, Polypeptides,
; FILE REFERENCE: PTO40P1
; CURRENT APPLICATION NUMBER: US/09/906,779
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US01/01563
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/176,306
; PRIOR FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-906-779-4

```

```

Query Match          16.7%; Score 76.5; DB 10; Length 603;
Best Local Similarity 27.0%; Pred. No. 3.8;
Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;

```

```

QY 7 PVSIVRAIQANINI-PMG-----AFRPGAGQPPRRKCTPEVEBGVPTSDDEKKPIPG 59
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 461 PISIQATIAKLSIRPFGLESFVASLPGPAEP-----GLPPASLPSTPIPS 509
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 60 AKKLPGPAVNLSEIQNKSEIKYVPAEQ 88
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 510 SSPPP-----LSSPLPEAPQPK 527

```

RESULT 3

```

US-10-020-215-2
; Sequence 2; Application US/10020215
; Publication No. US20030008347A1
; GENERAL INFORMATION:
; APPLICANT: PLOMMAN, GREGORY
; APPLICANT: PILES, EIOR
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF ALP RELATED DISORDERS
; FILE REFERENCE: 038602/1290
; CURRENT APPLICATION NUMBER: US/10/020,215
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 09/095,443
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/049,477

```

```

; PRIOR FILING DATE: 1997-06-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1274
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: ALP
; OTHER INFORMATION: polypeptide sequence
US-10-020-215-2

```

```

Query Match          16.7%; Score 76.5; DB 9; Length 1274;
Best Local Similarity 27.0%; Pred. No. 9.4;
Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;

```

```

QY 7 PVSIVRAIQANINI-PMG-----AFRPGAGQPPRRKCTPEVEBGVPTSDDEKKPIPG 59
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 1132 PISIQATIAKLSIRPFGLESFVASLPGPAEP-----GLPPASLPSTPIPS 1180
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 60 AKKLPGPAVNLSEIQNKSEIKYVPAEQ 88
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 1181 SSPPP-----LSSPLPEAPQPK 1198

```

RESULT 4

```

US-09-977-260-8
; Sequence 8; Application US/09977260
; Publication No. US20020192790A1
; GENERAL INFORMATION:
; APPLICANT: ULIRICH, AXEL
; APPLICANT: GISHITZY, MIKHAIL
; APPLICANT: SURES, IRMINGARD
; TITLE OF INVENTION: NOVEL MEGAKARYOCYTIC PROTEIN TYROSINE KINASES
; FILE REFERENCE: 038602/1260
; CURRENT APPLICATION NUMBER: US/09/977,260
; CURRENT FILING DATE: 2001-10-16
; PRIOR APPLICATION NUMBER: 08/232,545
; PRIOR FILING DATE: 1994-04-22
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-977-260-8

```

```

Query Match          16.0%; Score 73.5; DB 9; Length 659;
Best Local Similarity 32.4%; Pred. No. 8.7;
Matches 22; Conservative 13; Mismatches 14; Indels 19; Gaps 5;

```

```

QY 23 GARFPGAGQPPRRKCTPEVEBGVPTSDDE---KKRIPGAKKLPGPA---VNLSEIQNI 76
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 173 GSKLPSS---HRTKKP-----LPTPEEDQILKKVLP-----PEPAAPVSTSELKV 219
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
QY 77 KSELKTYV 84
   ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
DB 220 VALYDVP 227

```

RESULT 5

```

US-10-045-202-2
; Sequence 2; Application US/10045202
; Publication No. US20030040461A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb
; APPLICANT: Bristol-Myers Squibb
; TITLE OF INVENTION: MODULATORS OF BRUTON'S TYROSINE KINASE AND BRUTON'S TYROSINE KIN
; TITLE OF INVENTION: INTERMEDIATES AND METHODS FOR THEIR IDENTIFICATION AND USE IN T
; FILE REFERENCE: D0032 NP
; CURRENT APPLICATION NUMBER: US/10/045,202
; CURRENT FILING DATE: 2002-06-27

```

Sequence 5456, Application US/09815242
 Patent No. US20020061565A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard

```
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5456
; LENGTH: 978
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5456

Query Match          14.9%; Score 68.5; DB 10; Length 978;
Best Local Similarity 31.0%; Pred. No. 47;
Matches 26; Conservative 7; Mismatches 16; Indels 35; Gaps 5;

QY 27 PGAGPPRRKCTPEV---EEGVPTSD---EKKPIPGAKLP-----GP 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 864 PGEPTPP-----TPEVPSSEPTPTPTPEVPAPGKFPVPAKPEPKKSPVGGKVT 918
    : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 AVNLSE-----IONIKSEL 80
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 919 VIEINEKVAVAPTQKQSKSEL 942

RESULT 10
US-09-815-242-12686
; Sequence 12686, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12686
; LENGTH: 1001
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12686

Query Match          14.9%; Score 68.5; DB 10; Length 1001;
Best Local Similarity 31.0%; Pred. No. 48;
Matches 26; Conservative 7; Mismatches 16; Indels 35; Gaps 5;

QY 27 PGAGPPRRKCTPEV---EEGVPTSD---EKKPIPGAKLP-----GP 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 887 PGEPTPP-----TPEVPSSEPTPTPTPEVPAPGKFPVPAKPEPKKSPVGGKVT 941
    : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 AVNLSE-----IONIKSEL 80
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 942 VIEINEKVAVAPTQKQSKSEL 965

RESULT 11
US-09-858-664A-16
; Sequence 16, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00927-CIP
; CURRENT APPLICATION NUMBER: US/09/858,664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-16

Query Match          14.8%; Score 68; DB 10; Length 196;
Best Local Similarity 32.4%; Pred. No. 7.3;
Matches 24; Conservative 5; Mismatches 35; Indels 10; Gaps 3;

QY 8 VSNVRAIQANINIPMGAFRPGAGPP--RRKCTPEVEEGVPTSD---EKKPIPGA 60
    ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
DB 43 VSSAASSQADADKMSGTSTGSPSLPPGAPFAGSPASRRPPGADAEGRERAEPIPKM 102
    : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 KKLPGP---AVNLIS 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 103 KVLESFPRGAANAS 116

RESULT 12
US-09-981-353-95
; Sequence 95, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 95
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

```
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 2054053CD1
US-09-981-353-95

Query Match
Best Local Similarity 14.8%; Score 68; DB 9; Length 247;
Matches 30; Conservative 14; Mismatches 30; Indels 20; Gaps 7;

OY 4 SKOP-VSNVRAIQANINI-PMGAFRPGAGOPPRKRECTPEVE--EGVP-----PTSDEEK 55
DB 126 SKTPLVARKKVPFASVALPTAPSRGSGVQTPDELSSEVDAAEGAPVVGPKSEVEE- 184
OY 56 PIPGAKKLG-PA-----VNLSIQNIKSEIK 81
DB 185 ---GHGKLPSMPAAEMHAKVDELDQVIREIK 215

RESULT 13
US-09-925-300-1411
Sequence 1411, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Craig Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925.300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1411
LENGTH: 280
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-300-1411

Query Match
Best Local Similarity 14.8%; Score 68; DB 10; Length 280;
Matches 30; Conservative 14; Mismatches 30; Indels 20; Gaps 7;

OY 4 SKOP-VSNVRAIQANINI-PMGAFRPGAGOPPRKRECTPEVE--EGVP-----PTSDEEK 55
DB 159 SKTPLVARKKVPFASVALPTAPSRGSGVQTPDELSSEVDAAEGAPVVGPKSEVEE- 217
OY 56 PIPGAKKLG-PA-----VNLSIQNIKSEIK 81
DB 218 ---GHGKLPSMPAAEMHAKVDELDQVIREIK 248

RESULT 14
US-10-109-886-8
Sequence 8, Application US/10109886
Patent No. US20020119499A1
GENERAL INFORMATION:
APPLICANT: TANABE SEIYAKU CO. LTD.
APPLICANT: TANIGUCHI, Tomoyasu
APPLICANT: MIZUKAMI, Junko
TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND
FILE REFERENCE: TANIGUCHI-6
CURRENT APPLICATION NUMBER: US/10/109,886
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: 09/514,247
PRIOR FILING DATE: 2000-02-28
PRIOR APPLICATION NUMBER: PCT/JP98/03734
PRIOR FILING DATE: 1998-08-24
PRIOR APPLICATION NUMBER: JP231084/1997
PRIOR FILING DATE: 1997-08-27
NUMBER OF SEQ ID NOS: 10
```

```
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 2441
TYPE: PRT
ORGANISM: mouse
US-10-109-886-8

Query Match
Best Local Similarity 14.7%; Score 67.5; DB 12; Length 2441;
Matches 24; Conservative 14; Mismatches 35; Indels 23; Gaps 4;

OY 15 QANINPMGAFRPGAGOP-----PRKRECTPEVEEGVPPTSDEEKXPIPGAKKL 63
DB 1975 RANINNMPPGRGOMKGTTPGQMTPTVGLNVPNPVQSGPVMSMP-GQMQQAPIPQOQPM 2033
OY 64 PG---PAVNLS-----SIQNIKSEIKVPRKAO 88
DB 2034 PQMPRFVMSMQAAVAAGPMPVQGRNRSISPSALQ 2069

RESULT 15
US-09-864-761-45128
Sequence 45128, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
SEQ ID NO 45128
```

LENGTH: 180
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC019278.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.56
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.64
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.63
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.63
OTHER INFORMATION: EST_HUMAN HIT: BE964946.2, EVALUE 1.20e-01
OTHER INFORMATION: SWISSPROT HIT: Q05859, EVALUE 1.00e-68
US-09-647-761-45128

Query Match 14.6%; Score 67; DB 10; Length 180;
Best Local Similarity 28.2%; Pred. No. 8.4;
Matches 24; Conservative 12; Mismatches 29; Indels 20; Gaps 5;

QY 18 INIPMGAFRPGAGOP--PRRKE--CTPEVEE---GVPTSDEKKP-----IPGAKKL 63
DB 71 INIDMPKTEPKGADPESPRREMGCGNADQESQSGVGVQTOGGEVVKPSPETALEAFKAL 130
QY 64 -----PGPAVNLSEIQNIKSEIKY 82
DB 131 FIRPPRKGTADTSELKALKRKRMRH 155

Search completed: April 9, 2003, 12:53:40
Job time : 20.5161 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:29:50 ; Search time 31.5806 Seconds
(without alignments)
371.305 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459

Sequence: 1 MNMSKQPVSNVRAIQANINI.....NLSEIQIKSELKTVKPAQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	459	100.0	88	20	AA128651 Human chisel (Csl)
2	451	98.3	88	23	AAE16632 Human 66214 protei
3	382.5	83.3	85	20	AA128650 Murine chisel (Csl)
4	274	59.7	75	20	AA128634 Xenopus chisel (Cs
5	90	19.6	143	22	AA128495 Novel human secret
6	84	18.3	523	20	AA121855 Human signal pepti
7	78.5	17.1	4412	21	AA121855 Sequence g1/01742
8	77	16.8	1098	22	ABG20365 Novel human diagno
9	76.5	16.7	603	22	AA104485 Human protein cyro
10	76.5	16.7	691	22	AB803068 Human expressed po

11	76.5	16.7	691	22	AA123180
12	76.5	16.7	1253	21	AA129663
13	76.5	16.7	1264	21	AA129664
14	76.5	16.7	1274	20	AA129665
15	76.5	16.7	1583	21	AA129666
16	76.5	16.7	1636	21	AA129667
17	74	16.1	249	22	ABG27048
18	74	16.1	870	22	AA129668
19	73.5	16.0	191	21	AA129669
20	73.5	16.0	635	22	AA129670
21	73.5	16.0	659	17	AA129671
22	73.5	16.0	659	23	AA129672
23	72.5	15.8	1178	22	AA129673
24	72.5	15.8	2153	22	AA129674
25	72	15.7	322	21	AA129675
26	72	15.7	392	12	AA129676
27	72	15.7	428	12	AA129677
28	71.5	15.6	515	22	AA129678
29	71	15.5	91	22	AA129679
30	71	15.5	131	21	AA129680
31	71	15.5	874	22	AA129681
32	71	15.5	3502	22	AA129682
33	70.5	15.4	257	17	AA129683
34	70	15.3	146	22	AA129684
35	70	15.3	151	21	AA129685
36	70	15.3	172	21	AA129686
37	70	15.3	654	21	AA129687
38	69.5	15.1	197	22	AA129688
39	69.5	15.1	202	21	AA129689
40	69.5	15.1	452	20	AA129690
41	69.5	15.1	452	22	AA129691
42	69.5	15.1	452	22	AA129692
43	69.5	15.1	452	22	AA129693
44	69.5	15.1	452	22	AA129694
45	69.5	15.0	452	22	AA129695

ALIGNMENTS

RESULT 1
AA128651
ID AA128651 standard; Protein; 88 AA.
XX
AC AA128651;
DT 17-JAN-2000 (first entry)
DE Human chisel (Csl) protein.
XX
KW Chisel protein; Csl; EF-Hand protein super family; muscle development;
KW heart/skeletal muscle cell development; signalling pathway; regulation;
KW Xq21.3-q22; adaptive process; muscle homeostasis; cardiac hypertrophy;
KW detection; diagnosis; prophylaxis; treatment; differentiation; nucleus;
KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
KW skeletal myopathy; Becker's myotonic dystrophy; heart failure;
KW transgenic animal; drug screening; gene therapy; homology;
KW scallop regulatory myosin light chain.
XX
OS Homo sapiens.
XX
PN WO950410-A1.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-AU00220.
XX
PR 27-MAR-1998; 98AU-0002634.
XX
PA (CHAN-) CHANG CARDIAC RES INST VICTOR.
PA (GEO) GEN HOSPITAL CORP.
XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX

Novel human enzyme
Human tyrosine pho
Human tyrosine pho
Human ALP. Homo s
Human tyrosine pho
Human histidine do
Human human diagno
Human protein segn
Arabidopsis thalia
Human Bruton's tyr
BTK tyrosine kinase
Human Bruton's tyr
Drosophila melanog
Novel human secret
Arabidopsis thalia
Toxoplasma gondii
Toxoplasma gondii
Drosophila melanog
Human ORF ORF1406
Drosophila melanog
Drosophila melanog
HCV Toledo strain
Human polypeptide
Arabidopsis thalia
Arabidopsis thalia
Cks-P66-Cks fusion
Human colon cancer
Arabidopsis thalia
Polypeptide fragme
Human secreted pro
Human protein segn
Human protein segn
Human protein segn
Novel human secret

PI Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX WPI; 1999-610852/52.
 DR N-PSDB; AAX30904.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 7; Fig 3; 157bp; English.
 XX
 CC The present sequence is the human chisel protein (Csl), that is a member
 CC of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and has 86% homology
 CC to the mouse Csl. Structural homology between Csl and scallop regulatory
 CC myosin light chain is also detected. It is predominantly expressed in
 CC heart and skeletal muscles and is activated after the differentiation of
 CC cells. Csl functions in regulation aspects of differentiation or
 CC adaptive processes that maintain muscle homeostasis. This sequence can
 CC be used in the detection, diagnosis, prophylactic and therapeutic
 CC treatment of diseases such as those involving aberrant muscle cell
 CC development and functional activity. It is also used in the treatment of
 CC muscular and myotonic dystrophies, skeletal myopathies such as Duchenne
 CC muscular dystrophy and Becker's myotonic dystrophy, heart failure,
 CC cardiac hypertrophy, myocarditis, myofiber atrophy, etc. The Csl gene
 CC sequence can also be used in gene therapy, for the production of
 CC transgenic animals and for drug screening.
 XX
 SQ Sequence 88 AA;
 Query Match 100.0%; Score 459; DB 20; Length 88;
 Best Local Similarity 100.0%; Pred. No. 1.5e-40;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNMSKQPSVSNVRAIQANINIPMGAFRPGAGOPPRRKECTPEVEGVPPTSDEKKPIPGA 60
 DB 1 MNMSKQPSVSNVRAIQANINIPMGAFRPGAGOPPRRKECTPEVEGVPPTSDEKKPIPGA 60
 QY 61 KKLPGPAVNLSEIQNIKSELYPKAEQ 88
 DB 61 KKLPGPAVNLSEIQNIKSELYPKAEQ 88
 DT 09-APR-2002 (first entry)
 XX
 DE Human 66214 protein.
 XX
 KM Human; congestive heart failure; dilative cardiomyopathy; sudden death;
 KM hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
 KM heart muscle disease; conduction disorder; coronary heart disease;
 KM systemic arterial hypertension; pulmonary hypertension; endocarditis;
 KM pulmonary heart disease; valvular heart disease; pericardial disease;
 KM congenital heart disease; gene therapy; syncope; transgenic animal;
 KM expressed sequence tag; EST; 66214 protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200192567-A2.
 XX
 PD 06-DEC-2001.
 XX
 PP 30-MAY-2001; 2001WO-EP06165.
 XX
 PR 30-MAY-2000; 2000US-207400P.
 XX
 PA (MED1-) MEDIGENE AG.
 XX

PI Bunk D, Reuner B, Beck J, Henkel T;
 XX WPI; 2002-122073/16.
 DR N-PSDB; AAD27216.
 XX
 PT Identifying a subject at risk for a heart disease e.g. congestive heart
 PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 PT the polypeptide expressed by genes abnormally expressed in heart tissue
 PT
 XX
 PS Claim 1a; Fig 9c; 154bp; English.
 XX
 CC The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present sequence is 66214 protein encoded by an
 CC expressed sequence tag (EST) DNA.
 XX
 SQ Sequence 88 AA;
 Query Match 98.3%; Score 451; DB 23; Length 88;
 Best Local Similarity 98.3%; Pred. No. 1.1e-39;
 Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNMSKQPSVSNVRAIQANINIPMGAFRPGAGOPPRRKECTPEVEGVPPTSDEKKPIPGA 60
 DB 1 MNMSKQPSVSNVRAIQANINIPMGAFRPGAGOPPRRKECTPEVEGVPPTSDEKKPIPGA 60
 QY 61 KKLPGPAVNLSEIQNIKSELYPKAEQ 88
 DB 61 KKLPGPAVNLSEIQNIKSELYPKAEQ 88
 DT 17-JAN-2000 (first entry)
 XX
 DE Murine chisel (Csl) protein.
 XX
 KM Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KM heart/skeletal muscle cell development; signalling pathway; murine;
 KM X chromosome; regulation; adaptive process; muscle homeostasis; nucleus;
 KM detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KM Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KM differentiation; gene therapy; transgenic animal; drug screening;
 KM scallop regulatory myosin light chain.
 XX
 OS Mus sp.
 XX
 PN WO9950410-A1.
 XX
 PD 07-OCT-1999.
 XX
 PP 26-MAR-1999; 99WO-AU00220.
 XX
 PR 27-MAR-1998; 98AU-0002634.
 XX

PA (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 DR WPI, 1999-610852/52.
 DR N-PSDB; AAX90903.
 XX
 XX Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 XX Claim 2, Fig 3, 157pp; English.
 XX
 XX The present sequence is the murine chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 XX Sequence 85 AA;
 SQ
 Query Match 83.3%; Score 382.5; DB 20; Length 85;
 Best Local Similarity 86.0%; Pred. No. 1,4e-32;
 Matches 74; Conservative 3; Mismatches 8; Indels 1; Gaps 1;
 QY 3 MSKQPSNVTAIQANINIPMGAFRPGAGOPPRRKECTPEVEEGVPTSD--EKKRIP 62
 DB 1 MSKQPSNVTAIQANINIPMGAFRPGAGOPPRRKECTPEVEEGVPTSD--EKKRIP 62
 QY 63 LRGPAVNLSEIQNKSEKTVPKAEQ 88
 DB 60 FPGPVNLSEIQNKSEKTVPKAEQ 85
 RESULT 4
 AA128834
 ID AAY28834 standard; Protein; 75 AA.
 AC AAY28834;
 XX
 DT 17-JUN-2000 (first entry)
 DE Xenopus chisel (Csl) protein.
 XX
 KM Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KM heart/skeletal muscle cell development; signalling pathway; xenopus;
 KM X chromosome; regulation; adaptive process; muscle homeostasis; nucleus;
 KM detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KM Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KM differentiation; gene therapy; transgenic animal; drug screening;
 KM scallop regulatory myosin light chain.
 XX
 OS Xenopus sp.
 XX
 PN WO950410-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 26-MAR-1999; 99MO-AU00220.

XX
 PR 27-MAR-1998; 98AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 DR WPI, 1999-610852/52.
 XX
 XX Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 XX Claim 11, Fig 3, 157pp; English.
 XX
 XX The present sequence is the xenopus chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 XX Sequence 75 AA;
 SQ
 Query Match 59.7%; Score 274; DB 20; Length 75;
 Best Local Similarity 75.7%; Pred. No. 2,7e-21;
 Matches 56; Conservative 6; Mismatches 8; Indels 4; Gaps 2;
 QY 3 MSKQPSNVTAIQANINIPMGAFRPGAGOPPRRKECTPEVEEGVPTSD--EKKRIP 58
 DB 1 MSKQPSNVTAIQANINIPMGAFRPGAGOPPRRKECTPEVEEGVPTSD--EKKRIP 60
 QY 59 GAKKLRGPAVNLSE 72
 DB 61 GAVKLRGPAVNLSE 74
 RESULT 5
 AAU32495
 ID AAU32495 standard; Protein; 143 AA.
 AC AAU32495;
 XX
 DT 18-DEC-2001 (first entry)
 DE Novel human secreted protein #2986.
 XX
 KM Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-611725/70.
 DR Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 XX Claim 20; Page 623; 765pp; English.
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation, to regulate haematopoiesis, and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33104 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SO Sequence 143 AA;
 Query Match 19.6%; Score 90; DB 22; Length 143;
 Best Local Similarity 32.9%; Pred. No. 0.089;
 Matches 26; Conservative 8; Mismatches 25; Indels 20; Gaps 2;
 QY 23 GAPPAGAGOPRR-----KECPVEVEGVPPPTSDDEKRPICAKKLPGPAVNLSE----- 72
 DB 41 GGGPGGAAGPFRAGSGSRCKPELHSGAPTSVCPPLKPGSGGPGSQVRSSEDXGSAF 100
 QY 73 -----IQNIKSELK 81
 DB 101 PVRASMSNDSFQSIKALE 119
 RESULT 6
 AAAY21855
 ID AAAY21855 standard; Protein; 523 AA.
 XX
 AC AAAY21855;
 XX
 XX 20-SEP-1999 (first entry)
 XX
 DE Human signal peptide-containing protein (SIGP) (clone ID 3125156).
 XX
 KW Signal-peptide containing protein; SIGP; human; cancer; immune response;
 KW adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma; AIDS;
 KW Addison's disease; adult respiratory distress syndrome; allergy; anemia;
 KW asthma; atherosclerosis; bronchitis; cholecystitis; Crohn's disease;
 KW ulcerative colitis; atopic dermatitis; glomerulonephritis; emphysema;
 KW diabetes mellitus; atrophic gastritis; glomerulonephritis; gout; trauma;
 KW Grave's Disease; hyperostosis; irritable bowel syndrome; infection;
 KW lupus erythematosus; multiple sclerosis; myasthenia gravis; inflammation;
 KW osteoarthritis; osteoporosis; pancreatitis; polymyositis; scleroderma;
 KW rheumatoid arthritis; Sjogren's syndrome; autoimmune thyroiditis.
 XX
 OS Homo sapiens.
 XX
 XX WO9933981-A2.
 XX
 XX 08-JUL-1999.
 XX
 XX 22-DEC-1998; 98WO-US27598.

XX 31-DEC-1997; 97US-0002485.
 XX
 XX (INCY-) INCYTE PHARM INC.
 XX
 XX Baughn MR, Corley NC, Guegler KJ, Hillman JL, Lal P;
 PI Sather SK, Shah P;
 XX WPI; 1999-430242/36.
 DR N-PSDB; AAX82090.
 XX
 PT Human signal-peptide containing protein coding sequences used to
 PT treat cancer and immune responses
 XX
 XX Claim 1; Page 88-89; 99pp; English.
 XX
 XX The invention provides human signal-peptide containing proteins (SIGP)
 CC (AAAY21841-855) and polynucleotides (AAX82076-90) encoding the proteins.
 CC A host cell containing a vector comprising SIGP DNA can be used to
 CC produce the SIGP protein. The SIGP protein can be used, in conjunction
 CC with a pharmaceutical carrier to treat or prevent a cancer. An antagonist
 CC of the SIGP protein can be used to treat or prevent a cancer or an
 CC immune response. The cancers that can be treated or prevented include
 CC sarcomas, adenocarcinomas, leukemias, lymphomas, melanomas,
 CC teratocarcinomas, myelomas and cancers of the adrenal gland, bladder,
 CC bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,
 CC gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary,
 CC pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen,
 CC testis, thymus, thyroid, and uterus. The immune responses that can be
 CC treated or prevented include, AIDS, Addison's disease, adult respiratory
 CC distress syndrome, allergies, anemia, asthma, atherosclerosis,
 CC bronchitis, cholecystitis, Crohn's disease, ulcerative colitis, atopic
 CC dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic
 CC gastritis, glomerulonephritis, Grave's disease, gout, hyperostosis,
 CC irritable bowel syndrome, lupus erythematosus, multiple sclerosis,
 CC myasthenia gravis, myocardial or pericardial inflammation, rheumatoid
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid
 CC arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,
 CC complications of cancer, infections, and trauma.
 XX
 SO Sequence 523 AA;
 Query Match 18.3%; Score 84; DB 20; Length 523;
 Best Local Similarity 32.8%; Pred. No. 1.6;
 Matches 22; Conservative 10; Mismatches 17; Indels 18; Gaps 3;
 QY 7 PVSNTVAIQNNINI-PMG-----AFPPGAGOPRRRKECTPEVEGVPPPTSDDEKRP 59
 DB 435 PISIQATIKLSIRPGGLESFVPSLPGRAPEP-----GLPPASLPSTRTIPS 483
 QY 60 AKKLPGP 66
 DB 484 SSQTPFP 490
 RESULT 7
 AAAY3666
 ID AAAY3666 standard; Protein; 4412 AA.
 XX
 AC AAAY3666;
 XX
 XX 22-FEB-2000 (first entry)
 XX
 DE Sequence gi/1017427/emb/CAA62189 from an alignment with protein 608.
 XX
 KW Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;
 KW bone development; gi/1017427/emb/CAA62189.
 XX
 OS Unidentified.
 XX
 XX WO9960164-A1.
 XX
 XX 25-NOV-1999.

XX 14-MAY-1999; 99WO-US11066.
 XX 15-MAY-1998; 98US-0085673.
 XX (OUAR-) OUARK BIOTECH INC.
 PA
 PI Elhat P, Mor O, Skalter R, Felnstein E, Faerman A;
 XX WPI; 2000-053304/04.
 DR
 XX Identification of stress induced genes for determining risk and
 PT preventing, treating or controlling osteoporosis
 PS
 XX Claim 32; Fig 6A-R; 308pp; English.
 CC The present sequence is obtained from a clustrel X alignment with
 CC protein 608. Protein 608 was identified using the method of the invention
 CC after subjecting rat osteoblasts to mechanical stress. Expression of the
 CC 608 gene was found to be upregulated by about 3-fold in cells subjected
 CC to mechanical strain. The specification describes a method for the
 CC identification of genes responsive to a specific mechanical stress. The
 CC method comprises applying the mechanical stress to an organism (tissue
 CC or cells comprising bone cells), isolating the specific cellular
 CC fractions and extracting mRNA from them, and differentially analyzing the
 CC mRNA in comparison with control samples. The method is used to identify
 CC genes whose expression is responsive to a specific stress. The identified
 CC genes are employed in determining risk associated with a physiological or
 CC disease state. The risk determination methods are used for testing a
 CC the method of the invention. These medications, or genes identified by
 CC controlling a physiological or disease state (especially osteoporosis or
 CC bone density or other factors causing or contributing to osteoporosis or
 CC its symptoms or other conditions involved in mechanical stress or its
 CC lack. The methods can also be used for advancing research or studies in
 CC bone development.
 XX
 SO Sequence 4412 AA;
 Query Match 17.1%; Score 78.5; DB 21; Length 4412;
 Best Local Similarity 33.9%; Pred. No. 62;
 Matches 19; Conservative 9; Mismatches 27; Indels 1; Gaps 1;
 QY 32 PPRKCECTP-EVEGCVPTSDDEKKIPGAKKLPDPAVNLSIGNIKSELKYVKA 86
 Db 3735 PPKKPEVTVPKVPKPKVPEKKVPVPPKPKPEVPTKVPKVPKVAVPEKKVBEA 3790
 RESULT 8
 ABG20365
 ID ABG20365 standard; Protein; 1098 AA.
 AC
 XX ABG20365;
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #20356.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS84552.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS
 XX Claim 20; SEQ ID No 50724; 103pp; English.
 CC
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SO Sequence 1098 AA;
 Query Match 16.8%; Score 77; DB 22; Length 1098;
 Best Local Similarity 33.8%; Pred. No. 19;
 Matches 25; Conservative 7; Mismatches 26; Indels 16; Gaps 4;
 QY 2 NMSKQPSVNVRAIQANINIPMGAFRP---GAGOPPRKCECTPEVEGCVPT-----SDSE 53
 Db 732 NKEKTRSMKMSBARAKNTPATARRRRRRATPPAG---AQQQAQNAAPATATANDXE 788
 QY 54 KKPIPAKALPGPA 67
 Db 789 KTPRP-----PAPA 797
 RESULT 9
 AAU04485
 ID AAU04485 standard; Protein; 603 AA.
 AC
 XX AAU04485;
 DT 26-SEP-2001 (first entry)
 XX
 DE Human protein tyrosine phosphatase (PTPase) polypeptide #1.
 XX
 KW Protein tyrosine phosphatase; PTPase; human; nervous system; dementia;
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; autism;
 KW Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania;
 KW cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease;
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo;
 KW diabetes mellitus; Sjogren's syndrome; asthma; atopic dermatitis;
 KW allogenic transplant rejection; pancreatic disorder; sarcoidosis;
 KW leukaemia; cancer; neoplasm; Burkitt's lymphoma;
 XX
 OS Homo sapiens.
 XX
 PN WO200153530-A1.

XX 26-JUL-2001.
 PD 17-JAN-2001; 2001WO-US01563.
 XX 18-JAN-2000; 2000US-0176306.
 PF (HUMA-) HUMAN GENOME SCI INC.
 XX Shl Y, Ruben SM;
 PA WPI; 2001-442265/47.
 PI N-PSDB; AAS08886.
 DR
 DR
 PT New human protein tyrosine phosphatase polypeptide useful for treatment
 PT of disorders including disorder of the nervous system -
 PS Claim 11; Page 283-285; 290pp; English.
 XX The sequence represents a human protein tyrosine phosphatase (PTPase)
 CC polypeptide. PTPase proteins and nucleic acids are useful for preventing,
 CC treating or ameliorating a medical condition. The medical conditions are
 CC disorders of the central and peripheral nervous system (e.g. Alzheimer's
 CC disease, Huntington's disease, Parkinson's disease, Tourette's syndrome,
 CC obsessive compulsive disorder, schizophrenia, mania, dementia and
 CC autism), cardiovascular disorders (e.g. atherosclerosis and myocarditis),
 CC autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis,
 CC ankylosing spondylitis, Addison's disease, Sjogren's syndrome, vitiligo,
 CC asthma, atopic dermatitis and diabetes mellitus), inflammatory disorders
 CC (e.g. pancreatitis, sarcoidosis and allogeneic transplant rejection) and
 CC cancers and neoplasms (e.g. leukemia, Burkitt's lymphoma and multiple
 CC myeloma). The presence or absence of a mutation in the nucleic acid
 CC encoding the protein allows for diagnosis of or susceptibility to a
 CC pathological condition. The sequences are also useful in screening for
 CC agonists and antagonists of PTPase activity.
 XX
 SQ Sequence 603 AA;
 Query Match 16.7%; Score 76.5; DB 22; Length 603;
 Best Local Similarity 27.0%; Pred. No. 11;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQANIN-PMG-----AFRPGAGPPRRKCTPEVEGVPPPTSDDEKKRIPG 59
 Db 461 PISIQATIAKLSIRPGLESVPASLPAPBP-----GLPPASLPSTPIPS 509
 Db 510 SSPP-----LSSPLPEAPQPK 527
 QY 60 AKKLGPANVLSIQIKSEIKYVKAQ 88
 Db 510 SSPP-----LSSPLPEAPQPK 527
 RESULT 10
 ID ABB03068 standard; Protein; 691 AA.
 AC ABB03068;
 XX
 DT 07-JAN-2002 (first entry)
 DE Human expressed polypeptide SEQ ID NO 41.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO20015167-A1.
 XX 02-AUG-2001.
 PD
 XX

PF 17-JAN-2001; 2001WO-US01319.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 26-SEP-2000; 2000US-0235484.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-465559/50.
 DR N-PSDB; AA199620.
 XX
 PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with polypeptides and for
 PT treating autoimmune diseases e.g., multiple sclerosis, rheumatoid
 PT arthritis -
 PS Claim 11; SEQ ID NO 41; 504pp + Sequence Listing; English.
 XX The invention relates to novel genes (AA199614-AA199654) and proteins
 CC (AB03062-AB03085) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIP0 at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 691 AA;
 Query Match 16.7%; Score 76.5; DB 22; Length 691;
 Best Local Similarity 27.0%; Pred. No. 13;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQANIN-PMG-----AFRPGAGPPRRKCTPEVEGVPPPTSDDEKKRIPG 59
 Db 549 PISIQATIAKLSIRPGLESVPASLPAPBP-----GLPPASLPSTPIPS 597
 QY 60 AKKLGPANVLSIQIKSEIKYVKAQ 88
 Db 598 SSPP-----LSSPLPEAPQPK 615
 RESULT 11
 ID AAU23180 standard; Protein; 691 AA.
 AC AAU23180;
 XX
 DT 17-DEC-2001 (first entry)
 DE Novel human enzyme polypeptide #266.
 XX
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW

KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX Homo sapiens.
OS
XX
PN MO200155301-A2.
PD
XX
XX
PF 17-JAN-2001; 2001MO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198173.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216667.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-465566/50.
 DR N-PSDB; AAS41050.
 XX Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX Claim 11, SEQ ID No 1176; 1180bp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU2915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 691 AA;
 SQ
 Query Match 16.7%; Score 76.5; DB 22; Length 691;
 Best Local Similarity 27.0%; Pred. No. 13;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQNNINI-PMG-----AFRPGAGOPPRKKECTPEVEGVPTSDSEKKPIPG 59
 DB 549 PISIQTIKTLKIRPGGLSPVASLPGPAEP-----GLPPASLPSTPIPS 597
 QY 60 AKKLPGPAVNLSEIQNKSELKYVVKAEQ 88
 DB 598 SSPPP-----LSSPLPEAPQPKK 615
 RESULT 12
 AAB29663
 ID AAB29663 standard; Protein; 1253 AA.
 XX
 AC AAB29663;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Human tyrosine phosphatase HD-PTP cKAL16 fragment.
 XX
 KW Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
 KW chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
 KW lung cancer; tumour; gene therapy; diagnosis; recombinant production;
 KW anticancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200063392-A1.
 XX
 PD 26-OCT-2000.
 XX
 PR 14-APR-2000; 2000WO-JP02455.
 PA

XX
 PR 16-APR-1999; 99JP-0108842.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Shimizu K;
 XX
 DR WPI; 2000-672740/65.
 DR N-PSDB; AAC81226.
 XX
 XX Human tyrosine phosphatase with oncosteric activity encoded by a gene
 PT frequently deleted in lung cancer, useful for treatment and diagnosis
 PT of tumors
 XX
 XX Example 1; Page 98-106; 134pp; Japanese.
 XX
 CC The invention relates to a novel human tyrosine phosphatase, histidine
 CC domain-protein tyrosine phosphatase (HD-PTP; AAB29661) and to human
 CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The
 CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
 CC found to be deleted in lung cancers, and is therefore thought to contain
 CC a tumour suppressor gene. The invention also relates to expression
 CC vectors and host cells containing human HD-PTP nucleic acids; the
 CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
 CC gene therapy compositions containing DNA encoding HD-PTP; diagnostic
 CC reagents containing HD-PTP oligonucleotides; antibodies specific for
 CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
 CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
 CC and antibodies may be used in the treatment, investigation and diagnosis
 CC of cancers, particularly those of the lung. The present sequence
 CC represents a human HD-PTP fragment encoded by the cKAL16 cDNA fragment.
 XX
 SQ Sequence 1253 AA;
 Query Match 16.7%; Score 76.5; DB 21; Length 1253;
 Best Local Similarity 27.0%; Pred. No. 25;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQNNINI-PMG-----AFRPGAGOPPRKKECTPEVEGVPTSDSEKKPIPG 59
 DB 1111 PISIQTIKTLKIRPGGLSPVASLPGPAEP-----GLPPASLPSTPIPS 1159
 QY 60 AKKLPGPAVNLSEIQNKSELKYVVKAEQ 88
 DB 1160 SSPPP-----LSSPLPEAPQPKK 1177
 RESULT 13
 AAB29664
 ID AAB29664 standard; Protein; 1264 AA.
 XX
 AC AAB29664;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Human tyrosine phosphatase HD-PTP cKAL11 fragment.
 XX
 KW Human; histidine domain-protein tyrosine phosphatase; HD-PTP;
 KW chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
 KW lung cancer; tumour; gene therapy; diagnosis; recombinant production;
 KW anticancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200063392-A1.
 XX
 PD 26-OCT-2000.
 XX
 PF 14-APR-2000; 2000WO-JP02455.
 XX
 PR 16-APR-1999; 99JP-0108842.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Shimizu K;
 XX DR WPI; 2000-672740/65.
 XX DR N-PSDB; AAC81227.
 XX PT Human tyrosine phosphatase with oncostatic activity encoded by a gene
 PT frequently deleted in lung cancer, useful for treatment and diagnosis
 of tumors
 XX PS Example 1; Page 108-117; 134pp; Japanese.
 XX CC The invention relates to a novel human tyrosine phosphatase, histidine
 CC domain-protein tyrosine phosphatase (HD-PTP; AAB29661) and to human
 CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The
 CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
 CC found to be deleted in lung cancers, and is therefore thought to contain
 CC a tumour suppressor gene. The invention also relates to expression
 CC vectors and host cells containing human HD-PTP nucleic acids; the
 CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
 CC gene therapy compositions containing DNA encoding HD-PTP; diagnostic
 CC reagents containing HD-PTP oligonucleotides; antibodies specific for
 CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
 CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
 CC of cancer, particularly those of the lung. The present sequence
 CC represents a human HD-PTP fragment encoded by the cKAL11 DNA fragment.
 CC
 SQ Sequence 1264 AA;
 Query Match 16.7%; Score 76.5; DB 21; Length 1264;
 Best Local Similarity 27.0%; Pred. No. 25;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQANINI-PMG-----AFRGAGPPRRKCTBEVGEVPTSDKKPIPG 59
 Db 1122 PISISQATIAKLSIRPGLESPPVSLPGPAEP-----GLPPASLPSESTPIPS 1170
 QY 60 AKKLPGAVNLSEIQNIKSELYKVPKAO 88
 Db 1171 SSPP-----LSPLPEAPQPK 1188
 RESULT 14
 AAW89253
 ID AAW89253 standard; Protein; 1274 AA.
 XX AC AAW89253;
 XX DT 10-MAR-1999 (first entry)
 XX DE Human ALP.
 XX KM PRP04; PRP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
 KM type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
 KM neurodegenerative disease; neuronal survival; Alzheimer's disease;
 KM Parkinson's disease; Huntington's disease.
 XX OS Homo sapiens.
 XX PN WO9849317-A2.
 XX PD 05-NOV-1998.
 XX PF 27-APR-1998; 98WO-US08439.
 XX PR 23-OCT-1997; 97US-0063595.
 XX PR 28-APR-1997; 97US-0044428.
 XX PR 20-MAY-1997; 97US-0047222.
 XX PR 11-JUN-1997; 97US-0049477.
 XX PR 11-JUN-1997; 97US-0049756.
 XX PR 18-JUN-1997; 97US-0049914.

PA (SUGEN-) SUGEN INC.
 XX PI App H, Clary D, Courtenidge SA, Hui TH, Jallat B;
 XX PI Markby D, Onrust S, Peles E, Plozman GD;
 XX DR WPI; 1999-009434/01.
 XX DR N-PSDB; AAV81748.
 XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
 PT useful for identifying specific modulators for treatment and
 PT prevention of cancer and neurodegenerative disease
 XX PS Claim 2; Page 160-164; 193pp; English.
 XX CC The present invention describes isolated, enriched or purified nucleic
 CC acids encoding PRP04, SAD, PRP05, PRP10, ALP and ALK-7 proteins. The
 CC present sequence represents human ALP, the above proteins, other than
 CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
 CC substances that modulate their activity (i.e. agonists and antagonists,
 CC including NBP) in vivo or in vitro. These substances are used to treat
 CC or prevent diseases associated with abnormal signal transduction
 CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
 CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
 CC serine/threonine kinase) are used to promote neuronal survival.
 CC particularly for treating Alzheimer's, Parkinson's or Huntington's
 CC diseases. Nucleic acid fragments of the polynucleotides encoding the
 CC proteins can be used as probes to identify and clone related sequences;
 CC to detect protein-encoded RNA; to generate transgenic animals and in
 CC gene therapy (optionally after mutation). Ab are used to determine the
 CC proteins.
 SQ Sequence 1274 AA;
 Query Match 16.7%; Score 76.5; DB 20; Length 1274;
 Best Local Similarity 27.0%; Pred. No. 26;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;
 QY 7 PVSNNRAIQANINI-PMG-----AFRGAGPPRRKCTBEVGEVPTSDKKPIPG 59
 Db 1132 PISISQATIAKLSIRPGLESPPVSLPGPAEP-----GLPPASLPSESTPIPS 1180
 QY 60 AKKLPGAVNLSEIQNIKSELYKVPKAO 88
 Db 1181 SSPP-----LSPLPEAPQPK 1198
 RESULT 15
 AAB29662
 ID AAB29662 standard; Protein; 1583 AA.
 XX AC AAB29662;
 XX DT 23-FEB-2001 (first entry)
 XX DE Human tyrosine phosphatase HD-PTP fragment.
 XX KM Human, histidine domain-protein tyrosine phosphatase; HD-PTP;
 KM chromosome 3p21.3; gene deletion; tumour suppressor; cytostatic;
 KM lung cancer; tumour; gene therapy; diagnosis; recombinant production;
 KM anticancer.
 XX OS Homo sapiens.
 XX PN WO200063392-A1.
 XX PD 26-OCT-2000.
 XX PF 14-APR-2000; 2000WO-JP02455.
 XX PR 16-APR-1999; 99JP-0108842.
 XX PA (KYOWA) KYOWA HAKKO KOGYO KK.

PI Shimizu K;
 XX
 DR WPI; 2000-672740/65.
 N-PSDB; AAC81225.

XX Human tyrosine phosphatase with oncogenic activity encoded by a gene
 PT frequently deleted in lung cancer, useful for treatment and diagnosis
 PT of tumors -

XX
 PS Example 2; Page 82-98; 134pp; Japanese.

XX The invention relates to a novel human tyrosine phosphatase, histidine
 CC domain-protein tyrosine phosphatase (HD-PTP; A829661) and to human
 CC HD-PTP nucleic acids (AAC81224, AAC81225, AAC81262, AAC81263). The
 CC HD-PTP gene is located on chromosome 3p21.3. This region is frequently
 CC found to be deleted in lung cancers, and is therefore thought to contain
 CC a tumour suppressor gene. The invention also relates to expression
 CC vectors and host cells containing human HD-PTP nucleic acids; the
 CC recombinant production of HD-PTP; anticancer drugs containing HD-PTP;
 CC gene therapy compositions containing DNA encoding HD-PTP; diagnostic
 CC reagents containing HD-PTP oligonucleotides; antibodies specific for
 CC HD-PTP; and an immunoassay method using HD-PTP-specific antibodies for
 CC use in cancer diagnosis and investigation. HD-PTP proteins, nucleic acids
 CC and antibodies may be used in the treatment, investigation and diagnosis
 CC of cancers, particularly those of the lung. The present sequence
 CC represents a substantial portion of the human HD-PTP protein.

XX
 SQ Sequence 1583 AA;

Query Match 16.7%; Score 76.5; DB 21; Length 1583;
 Best Local Similarity 27.0%; Pred. No. 32;
 Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;

OY 7 PVSNNRAIQANINI-PMG-----AFRPGAGQPPRKECTPEVEGVPTSDDEKKPIPG 59
 Db 1441 PISSTQATIAKLSIRPPGGLSPVSLPQPAEP-----GLPPASLPSTPIPS 1489
 OY 60 AKKLPGPAVNLSEIQIKSELKYVPKAEQ 88
 Db 1490 SSPP-----LSSPLPEAPQPK 1507

Search completed: April 9, 2003, 12:35:31
 Job time : 32.5806 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:31:56 ; Search time 45.0645 Seconds

(without alignments)
402.360 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459
Sequence: 1 MNNSKQPSVNRVRAIQANINI.....NLSEIQIKSELYKVPRAEQ 88

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	382.5	83.3	85	11	Q9ER98 mus musculus
2	364.5	79.4	85	11	Q925F0 rat mus musculus
3	338	73.6	91	13	Q90258 xenopus lae
4	85.5	18.6	1209	4	Q9Y4D3 Q9Y4D3 homo sapien
5	81	17.6	5085	11	Q9JRK6 rat mus musculus
6	80.5	17.5	496	11	Q8RI25 mus musculus
7	80.5	17.5	508	11	Q923E6 mus musculus
8	79	17.2	1241	12	Q66579 mus musculus
9	78.5	17.1	1050	5	Q9SVZ3 eastern equ
10	78.5	17.1	1242	12	Q9SVZ3 dicyostell
11	78.5	17.1	7962	4	Q9PZX7 Q9PZX7 eastern equ
12	78.5	17.1	34350	4	Q10465 Q10465 homo sapien
13	78	17.0	182	2	Q8WZ42 Q8WZ42 homo sapien
14	78	17.0	1242	12	Q9PZX3 Q9PZX3 eastern equ
15	77.5	16.9	1242	12	Q88790 Q88790 eastern equ
16	77.5	16.9	1242	12	Q9PZX6 Q9PZX6 eastern equ

17	77.5	16.9	1242	12	Q9PZX5 Q9PZX5 eastern equ
18	77.5	16.9	1242	12	Q9PZX4 Q9PZX4 eastern equ
19	77.5	16.9	1242	12	Q9PZX2 Q9PZX2 eastern equ
20	77.5	16.9	1242	12	Q9PZX1 Q9PZX1 eastern equ
21	77.5	16.9	1242	12	Q9PZX0 Q9PZX0 eastern equ
22	77.5	16.9	1242	12	Q9PZW9 Q9PZW9 eastern equ
23	77.5	16.9	1242	12	Q88792 Q88792 eastern equ
24	77.5	16.9	1242	12	Q88793 Q88793 eastern equ
25	77.5	16.9	1242	12	Q88794 Q88794 eastern equ
26	77.5	16.9	1242	12	Q88795 Q88795 eastern equ
27	77.5	16.9	1242	12	Q88796 Q88796 eastern equ
28	77.5	16.9	1242	12	Q88797 Q88797 eastern equ
29	77.5	16.9	1242	12	Q88798 Q88798 eastern equ
30	77.5	16.9	1242	12	Q88799 Q88799 eastern equ
31	77.5	16.9	1242	12	Q08359 Q08359 eastern equ
32	77	16.8	458	10	Q24584 Q24584 zea mays (m
33	77	16.8	479	10	P93646 P93646 zea mays (m
34	76.5	16.7	160	2	Q93599 Q93599 bacteroides
35	76.5	16.7	577	4	Q9BSR5 Q9BSR5 homo sapien
36	76.5	16.7	990	4	Q9UG03 Q9UG03 homo sapien
37	76.5	16.7	1636	4	Q9H3S7 Q9H3S7 homo sapien
38	76	16.6	294	16	Q8YWC6 Q8YWC6 anabaena sp
39	75.5	16.6	3534	12	Q39266 Q39266 equine hevp
40	75.5	16.4	297	12	Q8V719 Q8V719 tt virus. o
41	74.5	16.2	1171	5	Q61732 Q61732 drosophila
42	73.5	16.0	95	4	Q9POL4 Q9POL4 homo sapien
43	73.5	16.0	676	6	Q95JC9 Q95JC9 sus scrofa
44	73	15.9	454	4	Q8WZ50 Q8WZ50 homo sapien
45	72.5	15.8	483	3	P79025 P79025 emeticelella

ALIGNMENTS

RESULT 1

Q9ER98 ID Q9ER98 PRELIMINARY; PRT; 85 AA.
AC Q9ER98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Stretch responsive muscle (X-chromosome) (SMX protein)
GN (Muscle-specific protein CSL).
OS SMX OR SRMX OR CSL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SKETELAL MUSCLE;
RX MEDLINE=2125047; PubMed=11401441;
RA Kemp T.V., Sadusky T.J., Simon M., Brown R., Eastwood R.,
RT "Identification of a Novel Stretch-Responsive Skeletal Muscle Gene (Smx).";
RL Genomics 72:260-271 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065879; PubMed=10598820;
RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.;
RT "Identification, mapping, and genomic structure of a novel X-chromosomal human gene (SMX) encoding a small muscular protein.";
RL Hum. Genet. 105:506-512 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Patzak D.;
RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=2125706; PubMed=11381084;
RA Palmer S., Groves N., Schindeler A., Yeoh T., Biben C., Wang C.-C.,

RA Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,
 RA Mohun T., Harvey R.P.;
 RT "The Small Muscle-specific Protein Csl Modifies Cell Shape and
 RT Promotes Myocyte Fusion in an Insulin-like Growth Factor 1-dependent
 RT Manner.";
 RL J. Cell Biol. 153:985-998(2001).
 DR EMBL; AJ245772; CAC08483.1; -
 DR EMBL; AF364070; AAK50388.1; -
 DR EMBL; AY026524; AAK07682.1; -
 DR MGI; MGI:191355; Smpx.
 SO SEQUENCE 85 AA; 9253 MW; 43863840A65DA6BC CRC64;

Query Match 83.3%; Score 382.5; DB 11; Length 85;
 Best Local Similarity 86.0%; Pred. No. 4.6e-34;
 Matches 74; Conservative 3; Mismatches 8; Indels 1; Gaps 1;

QY 3 MSKOPVSNVRAIQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 62
 DB 1 MSKOPISNVRALQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 59
 QY 63 LPGAVALSEIQNIKSELKYVPKAO 88
 DB 60 FPGPVNLSEIQNVKSELKFPKGEQ 85

RESULT 2

ID Q925F0 PRELIMINARY; PRT; 85 AA.
 AC Q925F0;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE SMPX protein.
 GN SMPX.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20065879; PubMed=10598820;
 RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.;
 RT "Identification, mapping, and genomic structure of a novel X-
 RT chromosome human gene (SMPX) encoding a small muscular protein.";
 RL Hum. Gene. 105:506-512(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Patzak D.;
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF364071; AAK50399.1; -
 SO SEQUENCE 85 AA; 9120 MW; 52654F8C790C932C CRC64;

Query Match 79.4%; Score 364.5; DB 11; Length 85;
 Best Local Similarity 81.4%; Pred. No. 4.1e-32;
 Matches 70; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 3 MSKOPVSNVRAIQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 62
 DB 1 MSKOPISNVRALQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 59
 QY 63 LPGAVALSEIQNIKSELKYVPKAO 88
 DB 60 FPGPVNLSEIQNVKSELKFPKGEQ 85

RESULT 3

ID Q90258 PRELIMINARY; PRT; 91 AA.
 AC Q90258;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Chisel.

GN CSL.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 CX NCBI_TaxID=8355;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=21275706; PubMed=11381084;
 RA Palmer S., Groves N., Schindeler A., Yeoh T., Biben C., Wang C.-C.,
 RA Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,
 RA Mohun T.J., Harvey R.P.;
 RT "The Small Muscle-specific Protein Csl Modifies Cell Shape and
 RT Promotes Myocyte Fusion in an Insulin-like Growth Factor 1-dependent
 RT Manner.";
 RL J. Cell Biol. 153:985-998(2001).
 DR EMBL; AF343894; AAK71068.1; -
 SO SEQUENCE 91 AA; 10006 MW; BD2BA90B82B3846C CRC64;

Query Match 73.6%; Score 338; DB 13; Length 91;
 Best Local Similarity 75.6%; Pred. No. 3.3e-29;
 Matches 68; Conservative 9; Mismatches 9; Indels 4; Gaps 2;

QY 3 MSKOPVSNVRAIQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 58
 DB 1 MSKOPASNIRIQANINIPMGAFRPGAGOPPRKCTPEVEEGVPTSDDEKKPIPGAK 60
 QY 59 GAKLPGPAVALSEIQNIKSELKYVPKAO 88
 DB 61 GAVLPGPAVALSEIQNVKSELKFPKRAEE 90

RESULT 4

ID Q9Y4D3 PRELIMINARY; PRT; 1209 AA.
 AC Q9Y4D3;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE KIAA0649 protein.
 GN KIAA0649.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Tissue=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kocani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 DR EMBL; AB014549; BAA31624.1; -
 SO SEQUENCE 1209 AA; 127352 MW; 2C6C92DC35E53C92 CRC64;

Query Match 18.6%; Score 85.5; DB 4; Length 1209;
 Best Local Similarity 35.1%; Pred. No. 1.3;
 Matches 27; Conservative 6; Mismatches 33; Indels 11; Gaps 2;

QY 13 AIQANINIPMGAFRPGAG-----QPPR---KCTPEVEEGVPTSDDEKKPIPGAK 61
 DB 136 AIOETLKAKSGAOPGAGAGOPGAGOPRAAGGSRCKPPEAHGAPVALCPPLVPSG 195
 QY 62 KLPGAVALSEIQNIKS 78
 DB 196 GPGSQVSSKQGSAS 212

RESULT 5

ID Q9JKS6

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC006582; AAH06582.1; -.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR InterPro; IPR000242; Tyr_PP.
 DR Pfam; PF0102; Y_phosphatase; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 FT NON_TER
 SQ SEQUENCE 508 AA; 55561 MW; 6B13C359PDE21513 CRC64;
 Query Match 17.5%; Score 80.5; DB 11; Length 508;
 Best Local Similarity 27.0%; Pred. No. 1.7;
 Matches 24; Conservative 14; Mismatches 22; Indels 29; Gaps 4;
 Oy 7 PVSNNRAIQANINIPMG-----AFPGAGPPRRKCTPEVGVPTSDDEKPIRG 59
 Db 366 PISIQATRLAKISIRPLGLDSPAISLPGLVHP-----GLPPASLPSTPVPS 414
 Oy 60 AKKLPGPAVNLSEIQNIKSELKYVPKARQ 88
 Db 415 SSPPP-----LSSPPLEAPQPE 432
 RESULT 8
 066579 PRELIMINARY; PRT; 1241 AA.
 AC 066579;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Eastern Equine encephalomyelitis virus RNA sequence.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OC NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SSP. NORTH AMERICAN VARIANT;
 RA MEDLINE=91375524; PubMed=1896061;
 RA Volchkov V.E.; Volchkova V.A.; Nemesov S.V.;
 RT "Complete nucleotide sequence of the genomic RNA of eastern equine encephalomyelitis virus."
 RT Mol. Gen. Microbiol. Virusol. 5:8-15(1991).
 RL EMBL; X6135; CAA44845.1; -.
 DR HSSP; P03315; 1VCP.
 DR InterPro; IPR001836; Alpha_core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR000936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.
 DR Pfam; PF00944; Alpha_core; 1.
 DR Pfam; PF01589; Alpha_E1_glycop; 1.
 DR Pfam; PF00943; Alpha_E2_glycop; 1.
 DR Pfam; PF01563; Alpha_E3_glycop; 1.
 DR PRINTS; PR00798; TOGAVIRIN.
 FT CHAIN 1 260 C PROTEIN.
 FT CHAIN 261 323 E3 PROTEIN.
 FT CHAIN 324 743 E2 PROTEIN.
 FT CHAIN 744 800 6K PROTEIN.
 FT CHAIN 801 1241 E1 PROTEIN.
 SQ SEQUENCE 1241 AA; 137559 MW; B966BB7416F0882 CRC64;
 Query Match 17.2%; Score 79; DB 12; Length 1241;
 Best Local Similarity 27.9%; Pred. No. 6.7;
 Matches 17; Conservative 14; Mismatches 24; Indels 6; Gaps 1;

Oy 8 VSNVRAIQANINIPMGAFPGAGPPRRKCTPEV-----EEGVPTSDDEKPIRG 61
 Db 41 IEDIRSNINLTLKQRAVPAPGAPPAKRRKPPAPSLRRKKRPPPAKOKRRKPPGR 100
 Oy 62 K 62
 Db 101 Q 101
 RESULT 9
 095VZ3 PRELIMINARY; PRT; 1050 AA.
 AC 095VZ3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Carnil.
 OS Dictyostelium discoideum (slime mold).
 OS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OC NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21319145; PubMed=11425877;
 RA Jung G.; Remmeit K.; Wu X.; Volosky J.M.; Hammer J.A. III;
 RT "The Dictyostelium CARMIL protein links capping protein and the Arp2/3 complex to type I myosins through their SH3 domains."
 RL J. Cell Biol. 153:1479-1497(2001).
 DR EMBL; AF385524; AAK72255.1; -.
 DR InterPro; IPR001993; Mitoch_carrier.
 DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
 SQ SEQUENCE 1050 AA; 113886 MW; D6F18849DAAC692B CRC64;
 Query Match 17.1%; Score 78.5; DB 5; Length 1050;
 Best Local Similarity 32.7%; Pred. No. 6.3;
 Matches 33; Conservative 12; Mismatches 33; Indels 23; Gaps 7;
 Oy 6 QVSNVRAIQANINIPMGAF--PGAGPPRRKCTPEVE--EGVPTSDDEKPIRG 59
 Db 926 QVSKFGAKLSANSAVAEARINMGAGAPIRKVPAPPEPPEVPTTKDVTPLKSKPVVA 985
 Oy 60 AKKLPGPAVNLSEIQNIKSELKYVPKARQ 83
 Db 986 PRSTPTTSTPTKTPVKKPSGSPVGLSDAPESDSELTHV 1026
 RESULT 10
 09PZX7 PRELIMINARY; PRT; 1242 AA.
 AC 09PZX7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Structural polypeptide.
 OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis virus).
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OC NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA38-MASS;
 RA Brault A.C.; Powers A.M.; Kang W.; Tesh R.B.; Shope R.E.; Weaver S.C.;
 RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis viruses from North, Central and South America."
 RT Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF159550; AAF04792.1; -.
 DR HSSP; P03315; 1VCP.
 DR InterPro; IPR001836; Alpha_core.
 DR InterPro; IPR002548; Alpha_E1_glycop.
 DR InterPro; IPR000936; Alpha_E2_glycop.
 DR InterPro; IPR002533; Alpha_E3_glycop.
 DR InterPro; IPR000930; Togavirin.

DR Pfam: PF00944; Alpha core; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KW Polypeptide; 1242 AA; 137667 MW; DP23D3631ACES1A CRC64;

Query Match 17.1%; Score 78.5; DB 12; Length 1242;
 Best Local Similarity 27.4%; Pred. No. 7.6;
 Matches 17; Conservative 14; Mismatches 24; Indels 7; Gaps 1;

Qy 8 VSNVAIAQNIIPMGAPPGAGQPPRRKCTPEVEEG-----VPTSDKKPIPGA 60
 Db 41 IEDLRVSNVLTIKORAPNPAPGAPKPKKAPKAPKAPKAPKAPKAPKAPKAPK 100

Qy 61 KK 62
 Db 101 RQ 102

RESULT 11

ID 010465 PRELIMINARY; PRT; 7962 AA.

AC 010465.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Titin, skeletal muscle isoform (EC 2.7.1.-) (Connectin) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKLELAL MUSCLE;
 RX MEDLINE=96026330; Pubmed=7569978;
 RA Labelt S.; Kolmer B.;
 RT "Titin: giant proteins in charge of muscle ultrastructure and elasticity."
 RL Science 270:293-296(1995).
 CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY ONE TISSUE
 CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
 CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE KINASES.
 CC -1- SIMILARITY: BELONGS TO IMMUGLOBULIN SUPERFAMILY. CONTAINS 90 IMMUGLOBULIN C2-LIKE DOMAINS.
 DR HSP; X90569; CAA62189.1; -.
 DR HSP; P56276; ITLK.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003600; IG_Like.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR004168; PPAK_motif.
 DR Pfam: PF00047; IG; 59
 DR Pfam: PF02818; PPAK; 53.
 DR SMART; SM00408; IGC2; 43.
 DR SMART; SM00410; IG_Like; 15.
 KW Muscle protein; Cytoskeleton; Structural protein; Phosphorylation; Serine/threonine-protein kinase; Alternative splicing; Repeat;
 KM Immunoglobulin domain
 FT NON_TER 1
 FT DOMAIN 5618 7792 GLU/LYS/PRO/VAL-RICH.
 FT NON_TER 7962 7962
 SQ SEQUENCE 7962 AA; 883018 MW; B8524053CBADE58 CRC64;

Query Match 17.1%; Score 78.5; DB 4; Length 7962;
 Best Local Similarity 33.9%; Pred. No. 64;

Matches 19; Conservative 9; Mismatches 27; Indels 1; Gaps 1;

Qy 32 PPRRRCETP-EVEGCVPTSEKKPIPGAKLPGAVNISEIGNISELKYVPA 86
 Db 7285 PPKKREVPVAPKAPKREVPKVPVPPKPKREVPPTKVPKVPKVPKVPK 7340

RESULT 12

ID 08W242 PRELIMINARY; PRT; 34350 AA.

AC 08W242.
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Titin.
 GN TTN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20309627; Pubmed=10850961;
 RA Freiburg A.; Trombetta K.; Hell W.; Gazorla O.; Fougereuse F.;
 RA Centner T.; Kolmerer B.; Witt C.; Beckmann J.S.; Gregorio C.C.;
 RA Granzler H.; Labelt S.;
 RT "Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity."
 RL Circ. Res. 86:1114-1121(2000).
 RN (2)

RP SEQUENCE FROM N.A.
 RX MEDLINE=21573839; Pubmed=11717165;
 RA Bang M.L.; Centner T.; Fornoff F.; Geach A.J.; Gotthardt M.;
 RA McNabb M.; Witt C.C.; Labelt D.; Gregorio C.C.; Granzler H.;
 RA Labelt S.;
 RT "The complete gene sequence of titin, expression of an unusual -700 kDa titin isoform and its interaction with obscurin identify a novel Z-line to I-band linking system."
 RL Circ. Res. 89:1065-1072(2001).
 DR EMBL; AJ277892; CAD12456.1; -.
 DR InterPro: IPR000282; Cytok_receptor_2.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR000577; FG_Y_kin.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001092; HLN_basic.
 DR InterPro: IPR003599; IG.
 DR InterPro: IPR003598; IG_C2.
 DR InterPro: IPR003006; IG_MHC.
 DR InterPro: IPR003596; IG_V.
 DR InterPro: IPR002016; PetoXidase.
 DR InterPro: IPR004168; PPAK_motif.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00041; FN3; 132.
 DR Pfam: PF00047; IG; 146.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF02818; PPAK; 53.
 DR PRODOM; PD000001; Euk_kinase; 1.
 DR SMART; SM00060; FN3; 133.
 DR SMART; SM00409; IG; 167.
 DR SMART; SM00408; IGC2; 148.
 DR SMART; SM00406; IGV; 23.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00933; FG_Y_KINASES_1; UNKNOWN 1.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN 1.
 DR PROSITE; PS00290; IG_MHC; UNKNOWN 1.
 DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN 1.
 SQ SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;

Query Match 17.1%; Score 78.5; DB 4; Length 34350;

Best Local Similarity 33.9%; Pred. No. 3.4e+02;
Matches 19; Conservative 9; Mismatches 27; Indels 1; Gaps 1;

QY 32 PPRKECTP-EVEGVPTSDSEKPIGAKLPPANVLSIINKSLKYVKA 86
DB 11525 PPKKEVTPVPEAPKEVPEKVPBPVKKEVPTKVEVPKVAPEKKVPA 11580

RESULT 13

Q9ADV0 PRELIMINARY; PRT; 182 AA.
AC Q9ADV0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 19.7 kDa protein.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
immunodominant 30-kilodalton major outer membrane proteins of
Ehrlichia canis and application of the recombinant protein for
serodiagnosis.";
RT J. Clin. Microbiol. 36:2671-2680(1998).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
Chaffeensis.";
RT Infect. Immun. 69:2083-2091(2001).
RL EMBL; AF078553; AAK28702.1; -.
DR Hypothetical protein.
KW SEQUENCE 182 AA; 19749 MW; 71B4319122EAEF52 CRC64;

Query Match 17.0%; Score 78; DB 2; Length 182;
Best Local Similarity 31.1%; Pred. No. 0.97;
Matches 19; Conservative 11; Mismatches 21; Indels 10; Gaps 2;

QY 6 QPVSVRAIQANINIPM-----GAPRPGAGQPPRKECTPEVEGV--PPTDDEKK 55
DB 119 QPVTAKSQAEVSKVPTDVKENQDAPQAPANTPQEKQESVPTPADGVNDDPTKEGASQ 178

QY 56 P 56
DB 179 P 179

RESULT 14

Q9PZX3 PRELIMINARY; PRT; 1242 AA.
AC Q9PZX3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
virus)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FL93-939;
RA Brault, A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;

RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
viruses from North, Central and South America.";
RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF159554; AAP04796.1; -.

DR HSSP; P03315; 1VCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polypeptide.
SQ SEQUENCE 1242 AA; 137613 MW; 6F474E82A91PF4CD CRC64;

Query Match 17.0%; Score 78; DB 12; Length 1242;
Best Local Similarity 32.7%; Pred. No. 8.6;
Matches 18; Conservative 12; Mismatches 21; Indels 4; Gaps 1;

QY 8 VSNVRAIQANINIPMGAPRPGAGQPPRKECTPEVEGVPTSDSEKPIPGAKK 62
DB 41 IEDLRKSIANTLTKORAPNPAPGPAKKKPAKPKPKKPKKPKKPKKPKK 91

RESULT 15

Q88790 PRELIMINARY; PRT; 1242 AA.

AC Q88790;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Structural polypeptide.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
virus)
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NORTH AMERICAN ANTIGENIC VARIETY;
RX MEDLINE=94025587; PubMed=8105605;
RA Weaver S.C., Hagedaugh A., Bellow L., Netesov S.V., Volchokov V.E.,
Chang G.D., Clacke D.K., Goussert L., Scott T.W., Trent D.W.,
Holland J.J.;
RT "A comparison of the nucleotide sequences of eastern and western
equine encephalomyelitis viruses with those of other alphaviruses and
related RNA viruses.";
RT Virology 197:375-390(1993).
RL EMBL; U01034; AAC53735.1; -.

DR HSSP; P03315; 1VCP.
DR InterPro; IPR001836; Alpha_core.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR000930; Togavirin.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polypeptide.
FT CHAIN 1 261 CAPSID.
FT CHAIN 262 324 E3.
FT CHAIN 325 744 E2.
FT CHAIN 745 801 E1.
FT CHAIN 802 1242 E1.
SQ SEQUENCE 1242 AA; 137627 MW; BCA1B498B4AABD27 CRC64;

Query Match 17.0%; Score 78; DB 12; Length 1242;
Best Local Similarity 32.7%; Pred. No. 8.6;

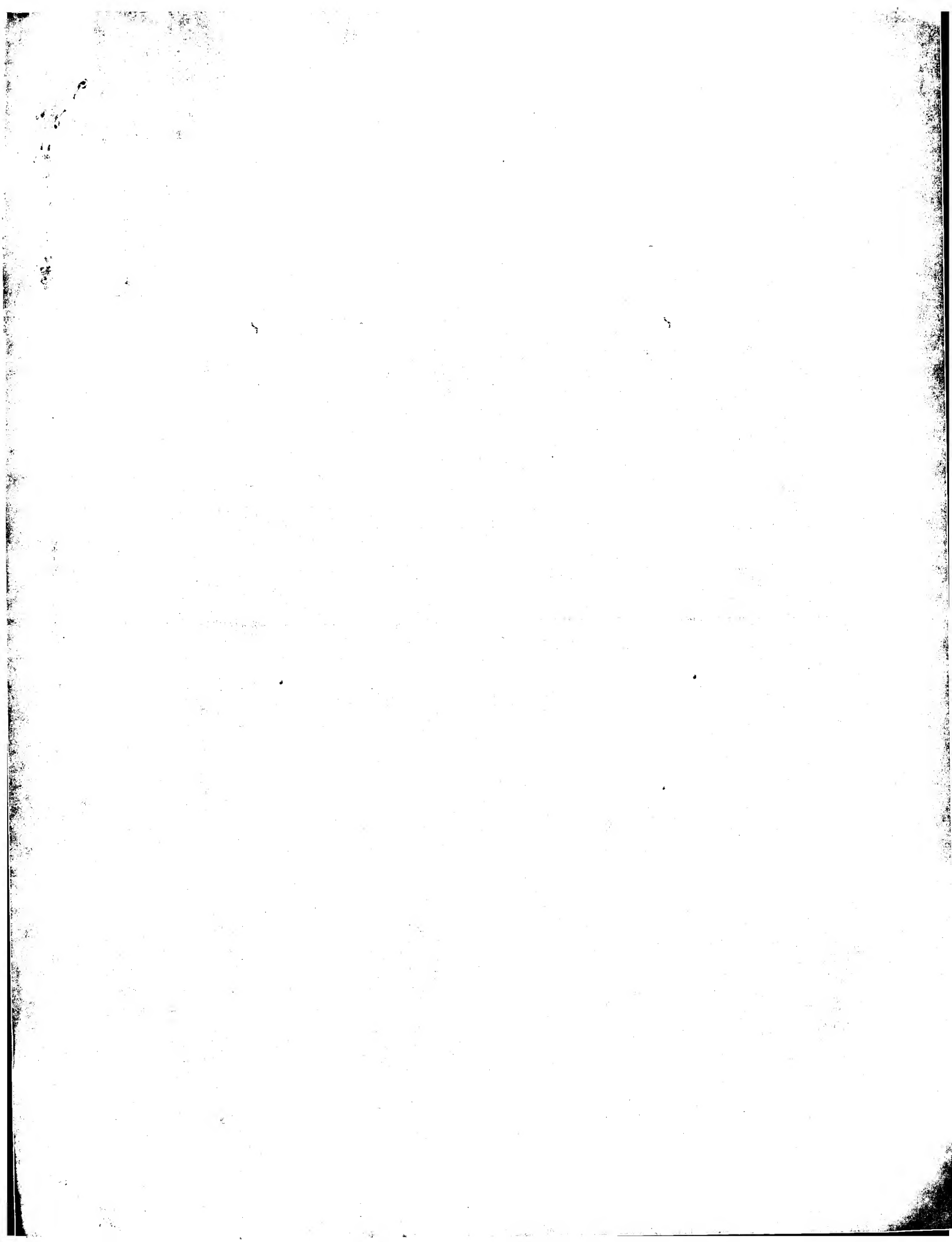
Wed Apr 16 05:43:23 2003

us-09-647-019-4.rpt

Page 7

	Matches	18;	Conservative	12;	Mismatches	21;	Indels	4;	Gaps	1;
Qy	8	VSNVRAIQANINIPMGAFRPGAGQPRRKCECTPEVEEGVPPTSDKKRTI	PAAKK	62	:	:	:	:	:	:
Dd	41	IEDLRRLNLTQLKORAPNPAGPPAKRKPKP----	PQTGKKRP	91	:	:	:	:	:	:

Search completed: April 9, 2003, 12:38:34
Job time : 54.0645 secs



GenCore version 5.1.4_p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 9, 2003, 12:31:10 ; Search time 12.0645 Seconds

(without alignments)
302.533 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459

Sequence: 1 MNMSKQPVSNVRAIQANINI.....NLSEIQNIKSELKTPVRAEQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	459	100.0	88	1 SMPX_HUMAN	Q9uh99 homo sapien
2	79	17.2	1239	1 POLS_EERY	P08768 eastern equ
3	77.5	16.9	1240	1 POLS_EERY3	P27284 eastern equ
4	76.5	16.7	757	1 HT16_HYDAT	P53356 hydra atten
5	75.5	16.4	1072	1 MAP2_BOVIN	P36225 bos taurus
6	74.5	16.2	416	1 NAPS_HUMAN	O14513 homo sapien
7	74	16.1	2142	1 BAT2_HUMAN	P48634 homo sapien
8	73.5	16.0	659	1 BTK_HUMAN	P06187 homo sapien
9	69.5	15.1	393	1 HXAA_HUMAN	P31260 homo sapien
10	69.5	15.1	825	1 SES_RAT	O63003 rattus norv
11	69	15.0	204	1 MSX3_MOUSE	P70354 mus musculu
12	69	15.0	874	1 BCAL_MOUSE	O61140 mus musculu
13	69	15.0	968	1 BCAL_RAT	O63767 rattus norv
14	69	15.0	1484	1 CES2_HUMAN	O9bx73 homo sapien
15	68.5	14.9	372	1 VGLI_HY2H	P13291 herpes simp
16	68	14.8	267	1 SSP2_HUMAN	P28290 homo sapien
17	68	14.8	407	1 ADRM_HUMAN	O16186 homo sapien
18	68	14.8	407	1 ADRM_MOUSE	O9jxv1 mus musculu
19	68	14.8	407	1 ADRM_RAT	O9jxv5 rattus norv
20	68	14.8	514	1 AMPL_YEAST	P14504 saccharomyc
21	68	14.8	3038	1 TRIO_HUMAN	O75962 homo sapien
22	67.5	14.7	831	1 NFH_RAT	P16884 rattus norv
23	67.5	14.7	2441	1 CBP_MOUSE	P45481 mus musculu
24	67	14.6	543	1 VP61_NRYAC	O03309 autographa
25	67	14.6	994	1 CLC1_MOUSE	O05623 gallus gall
26	66.5	14.5	537	1 MYPH_CHICK	P51125 mus musculu
27	66	14.4	788	1 ICAL_MOUSE	P56945 homo sapien
28	66	14.4	1073	1 BCAL_HUMAN	P53564 saccharomyc
29	65.5	14.3	381	1 YF04_YEAST	O58341 methanococc
30	65.5	14.3	511	1 TH11_MERTJA	P07948 homo sapien
31	65.5	14.3	511	1 LYN_HUMAN	O15240 homo sapien
32	65.5	14.3	616	1 VGF_HUMAN	P35991 mus musculu
33	65.5	14.3	659	1 BTK_MOUSE	

34	65.5	14.3	732	1 TAU_MOUSE	P10637 mus musculu
35	65.5	14.3	999	1 OXRP_CRIGR	O60432 cricetulus
36	65.5	14.3	1043	1 TCF8_MESAU	O60542 mesocricetu
37	65.5	14.3	1236	1 POLS_MESAU	P13897 western equ
38	65	14.2	774	1 NEK1_MOUSE	P51954 mus musculu
39	65	14.2	994	1 CLC1_RAT	P35524 rattus norv
40	65	14.2	999	1 OXRP_HUMAN	O9y411 homo sapien
41	65	14.2	1010	1 VINI_MOUSE	P19826 caenorhabdi
42	65	14.2	1132	1 BAT3_HUMAN	P46379 homo sapien
43	64.5	14.1	448	1 AAC2_DICDI	P14196 dictyosteli
44	64.5	14.1	533	1 LCP2_MOUSE	O60787 mus musculu
45	64.5	14.1	536	1 CP74_LINUS	P48417 linum usita

ALIGNMENTS

RESULT 1
ID SMPX_HUMAN STANDARD; PRT; 88 AA.

AC Q9UH99;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Small muscular protein (Stretch responsive skeletal muscle protein).
GN SMPX OR SRMX.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065679; PubMed=10598820;

RA Patzka D., Zhuchenko O., Lee C.-C., Wehnert M.;
RT "Identification, mapping, and genomic structure of a novel X-
chromosomal human gene (SMPX) encoding a small muscular protein.";
RT Hum. Genet. 105:506-512(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=skeletal muscle;
RA Kemp T.J., Sadusky T.J., Carey N., Coulton G.R.;
RT "Identification of a stretch responsive skeletal muscle gene.";
RN Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RP [3]
RC SEQUENCE FROM N.A.
RA TISSUE=skeletal muscle;
RA Strausberg R.;
RT Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY AND ABUNDANTLY EXPRESSED IN
CC HEART AND SKELETAL MUSCLE.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF129505; AAF19343.1;
CC EMBL; AJ250584; CAC08492.1;
CC EMBL; BC005948; AAH05948.1;
CC DR Gene; HGNC:11122; SMPX.
CC MIM: 300226;
CC SEQUENCE 88 AA; 9559 MW; CE33D2839F0F9EB7 CRC64;

Query Match 100.0%; Score 459; DB 1; Length 88;
Best local similarity 100.0%; Pred. No. 1.6e-35;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNMSKQPVSNVRAIQANINIPKGFAPGACQPPRRKCTREVEEGVPTSDDEKKPIPGA 60

Db 1 MMSKQVSNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPTSDKKPIPGA 60
 QY 61 KKLPGPAVNLSEIONIKSELKYVPAQ 88
 Db 61 KKLPGPAVNLSEIONIKSELKYVPAQ 88

RESULT 2
 ID POLS_EEEV STANDARD; PRT; 1239 AA.
 AC P08768;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [contains: Coat protein C (EC 3.4.21.-)
 DE (capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
 DE 6 kDa peptide; Spike glycoprotein E1].
 OS Eastern equine encephalitis virus [Eastern equine encephalomyelitis
 OS virus].
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OX NCBI_TaxID=11021;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=82V-2137;
 RX MEDLINE=87282265; PubMed=2886548;
 RA Chang G.-J., Trent D.W.;
 RT "Nucleotide sequence of the genome region encoding the 26S mRNA of
 RT eastern equine encephalomyelitis virus and the deduced amino acid
 RT sequence of the viral structural proteins.";
 RL J. Gen. Virol. 68:2129-2142(1987).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X05816; CAA29261.1; -
 CC PIR: A26816; VHWVEE.
 CC HSSP: P03315; 1VCP.
 DR MEROPS: S03.001; -
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF00944; Alpha_core; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KM Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
 KM Serine protease.
 FT CHAIN 1 259 COAT PROTEIN C.
 FT CHAIN 260 322 SPIKE GLYCOPROTEIN E3.
 FT CHAIN 323 742 SPIKE GLYCOPROTEIN E2.
 FT CHAIN 743 798 6 KDA PEPTIDE.
 FT CHAIN 799 1239 SPIKE GLYCOPROTEIN E1.
 FT ACT_SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRANSMEM 261 277 POTENTIAL.
 FT TRANSMEM 684 701 POTENTIAL.
 FT TRANSMEM 727 737 POTENTIAL.

FT TRANSMEM 777 798 POTENTIAL.
 FT TRANSMEM 1211 1235 POTENTIAL.
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 932 932 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1239 AA; 137431 MW; 8C7664A405D2D41C CRC64;
 Query Match 17.2%; Score 79; DB 1; Length 1239;
 Best Local Similarity 27.9%; Pred. No. 8.7;
 Matches 17; Conservative 14; Mismatches 24; Indels 6; Gaps 1;
 QY 8 VGNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPTSDKKPIPGA 61
 Db 40 IDLRSTANLTLTKRAPNPAPPAKKKAPSLSTETKKRPPPAKKKRPKPKR 99
 QY 62 K 62
 Db 100 Q 100

RESULT 3
 ID POLS_EEEV3 STANDARD; PRT; 1240 AA.
 AC P27284;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Structural polypeptide (P130) [contains: Coat protein C (EC 3.4.21.-)
 DE (capsid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
 DE 6 kDa peptide; Spike glycoprotein E1].
 OS Eastern equine encephalitis virus (strain va33[ten broeck]) (Eastern
 OS equine encephalomyelitis virus).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OX NCBI_TaxID=11022;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220727; PubMed=2024496;
 RA Weaver S.C., Scott T.W., Rico-Hesse R.;
 RT "Molecular evolution of eastern equine encephalomyelitis virus in
 RT North America.";
 RL Virology 182:774-784(1991).
 CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
 CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THE 6 kDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
 CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
 CC HEMAGGLUTININ.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M69094; AAA42980.1; -
 CC PIR: A39992; VHWVEV.
 CC HSSP: P03315; 1VCP.
 DR MEROPS: S03.001; -
 DR InterPro: IPR002548; Alpha_E1_glycop.
 DR InterPro: IPR000936; Alpha_E2_glycop.
 DR InterPro: IPR002533; Alpha_E3_glycop.
 DR InterPro: IPR001836; Alpha_core.
 DR InterPro: IPR000930; Togavirin.
 DR Pfam: PF00943; Alpha_E2_glycop; 1.
 DR Pfam: PF00944; Alpha_core; 1.
 DR Pfam: PF01563; Alpha_E3_glycop; 1.
 DR Pfam: PF01589; Alpha_E1_glycop; 1.
 DR PRINTS: PR00798; TOGAVIRIN.
 KM Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
 KM Serine protease.

FT CHAIN 1 260 COAT PROTEIN C.
 FT CHAIN 261 323 SPIKE GLYCOPROTEIN E3.
 FT CHAIN 324 343 SPIKE GLYCOPROTEIN E2.
 FT CHAIN 744 743 6 KDA PEPTIDE.
 FT CHAIN 799 1240 SPIKE GLYCOPROTEIN E1.
 FT ACT_SITE 137 137 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 211 211 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT TRANSMEM 259 276 POTENTIAL.
 FT TRANSMEM 695 712 POTENTIAL.
 FT TRANSMEM 722 738 POTENTIAL.
 FT TRANSMEM 781 799 POTENTIAL.
 FT TRANSMEM 1212 1236 POTENTIAL.
 FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 834 834 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 933 933 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1240 AA; 137290 MW; ABBEB159D083045 CRC64;

Query Match 16.9%; Score 77.5; DB 1; Length 1240;
 Best Local Similarity 27.4%; Pred. No. 12;
 Matches 17; Conservative 14; Mismatches 24; Indels 7; Gaps 1;

QY 8 VSNVRAIQANINIPMGARPGAGPPRRKCTPEVEEG-----VPTSEKKPIPGA 60
 Db 40 IEDLRSTIANLTKORAPNPAGPKKKPKPKPAQAKKQPPPAKKQKPKPKGK 99

QY 61 KK 62
 Db 100 RQ 101

RESULT 4
 HT16 HYDAT STANDARD; PRT; 757 AA.

AC P5356;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase HTK16 (EC 2.7.1.112).
 GN HTK16.
 OS Hydra attenuata (Hydra) (Hydra vulgaris).
 OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
 OC Hydrozoa; Hydra.
 ON NCBI_TaxID=6087;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Irvine;
 RX MEDLINE=94181280; PubMed=8134129;
 RA Chan T.A., Chu C.A., Rauen K.A., Krolier M., Tatarewicz S.M.,
 RA Steele R.E.;
 RT "Identification of a gene encoding a novel protein-tyrosine kinase
 RT containing SH2 domain and ankyrin-like repeats."
 RL Oncogene 9:1253-1259(1994).
 CC - FUNCTION: MAY BE INVOLVED IN SIGNAL TRANSDUCTION.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - TISSUE SPECIFICITY: EPITHELIAL CELLS.
 CC - SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
 CC - SIMILARITY: CONTAINS 2 SH2 DOMAINS.
 CC - SIMILARITY: CONTAINS 5 ANK REPEATS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; U00936; AAC27350.1; -
 CC HSSP; P08631; 1AD5
 DR InterPro; IPR002110; ANK.

DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00017; SH2_2.
 DR Pfam; PF00023; ank; 3.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR ProDom; PD000093; SH2_2.
 DR SMART; SM00248; ANK; 4.
 DR SMART; SM00252; SH2; 2.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50001; SH2_2.
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 FT SH2 domain; ANK repeat; Repeat.
 FT DOMAIN 10 102 SH2 1.
 FT REPEAT 115 147 ANK 1.
 FT REPEAT 151 180 ANK 2.
 FT REPEAT 184 214 ANK 3.
 FT REPEAT 219 248 ANK 4.
 FT REPEAT 252 281 ANK 5.
 FT DOMAIN 287 379 SH2 2.
 FT DOMAIN 484 740 PROTEIN KINASE.
 FT NP_BIND 490 498 ATP (BY SIMILARITY).
 FT BINDING 516 516 ATP (BY SIMILARITY).
 FT ACT_SITE 608 608 BY SIMILARITY.
 FT MOD_RES 746 746 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 757 AA; 85598 MW; 144D09E9109D794F CRC64;

Query Match 16.7%; Score 76.5; DB 1; Length 757;
 Best Local Similarity 33.3%; Pred. No. 8.8;
 Matches 25; Conservative 10; Mismatches 27; Indels 13; Gaps 3;

QY 2 NMKSQPSNVRAIQANINIPMGARPGAGPPRRKCTPEVEEGVPTSD-EKKPIPGA 60
 Db 396 NQKLTPVPPSRPKNNNGP-----QPLPYPTNESDSDITRLCEKPKLP-- 444

QY 61 KKLPGPAVNLSEIQN 75
 Db 445 -KLPRPVNHTTEVPN 458

RESULT 5
 MAP4 BOVIN STANDARD; PRT; 1072 AA.

AC P3625;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Microtubule-associated protein 4 (MAP 4) (Microtubule-associated
 DE protein-U) (MAP-U).
 GN MAP4.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=90338002; PubMed=2380192;
 RA Aizawa H., Emori Y., Murofushi H., Kawasaki H., Sakai H.,
 RA Suzuki K.;
 RT "Molecular cloning of a ubiquitously distributed microtubule-
 RT associated protein with Mr 190,000."
 RL J. Biol. Chem. 265:13849-13855(1990).
 RN [2]
 RP DOMAINS.
 RX MEDLINE=91236765; PubMed=2033072;

RA Aizawa H., Emori Y., Mori A., Murofuchi H., Sakai H., Suzuki K.;
 RT "Functional analyses of the domain structure of microtubule-associated
 RT protein-4 (MAP-4).";
 RL J. Biol. Chem. 266:9841-9846(1991).
 CC -1- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES
 CC MICROTUBULE ASSEMBLY.
 CC -1- TISSUE SPECIFICITY: IS DISTRIBUTED UBQUITOUSLY AMONG ALL TISSUES
 CC BUT AMOUNTS ARE LOWER IN CEREBELLUM AND LIVER.
 CC -1- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS
 CC NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE
 CC ASSEMBLY.
 CC -1- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; D90149; BAA14179.1; -.
 CC DR PIR; A37127; A37127.
 CC DR InterPro; IPR001084; Tubulin Tau.
 CC DR Pfam; PF00418; tubulin-binding; 3.
 CC DR PROSITE; PS00229; TAU MAP; 3.
 CC KM Microcubules; Repeat; Phosphorylation.
 CC FT DOMAIN 244 530 19 X 14 AA TANDEM REPEATS.
 CC FT REPEAT 244 257 1.
 CC FT REPEAT 258 271 2.
 CC FT REPEAT 272 285 3.
 CC FT REPEAT 286 299 4.
 CC FT REPEAT 300 313 5.
 CC FT REPEAT 314 327 6.
 CC FT REPEAT 328 341 7.
 CC FT REPEAT 342 355 8.
 CC FT REPEAT 384 391 9 (INCOMPLETE).
 CC FT REPEAT 392 405 10.
 CC FT REPEAT 406 417 11.
 CC FT REPEAT 418 431 12.
 CC FT REPEAT 432 445 13.
 CC FT REPEAT 446 460 14.
 CC FT REPEAT 461 474 15.
 CC FT REPEAT 475 488 16.
 CC FT REPEAT 489 502 17.
 CC FT REPEAT 503 516 18.
 CC FT REPEAT 517 530 19.
 CC FT REPEAT 907 937 TAU/MAP MOTIF 1.
 CC FT REPEAT 938 968 TAU/MAP MOTIF 2.
 CC FT REPEAT 969 1000 TAU/MAP MOTIF 3.
 CC SQ SEQUENCE 1072 AA; 111914 MW; E8C17A730989F0D2 CRC64;
 Query Match 16.4%; Score 75.5; DB 1; Length 1072;
 Best Local Similarity 27.4%; Pred. No. 16;
 Matches 23; Conservative 18; Mismatches 36; Indels 7; Gaps 4;
 QY 2 NMSKQPVSNVRAIQANINIPMGAFRPGAGOPRR-RKECTPEVERGCVPTSD-EKKKIP 58
 DB 804 DLSRPKSTTSVKKSTVPGTA--PPAGAPSRAPRTATPRPGSTPPVDKKPTAKPTS 861
 QY 59 GAKTLPGPAVNL--EIONIKSEL 80
 DB 862 SAPRLGRVANAASAPDLNVRSKV 885
 RESULT 6
 NAPS HUMAN
 ID NAPS HUMAN STANDARD; PRT; 416 AA.
 AC 014513;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nck-associated protein 5 (NAP-5) (Fragment).

GN NAPS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9800866; PubMed=9344857;
 RA Matoka K., Miki H., Takahashi K., Takenawa T.;
 RT "A novel ligand for an SH3 domain of the adapter protein Nck bears an
 RT SH2 domain and nuclear signaling motifs";
 RL Biochem. Biophys. Res. Commun. 239:488-492(1997).
 CC -1- SUBUNIT: Interacts with the SH3-containing region of the adapter
 CC protein Nck.
 CC -1- TISSUE SPECIFICITY: Expressed in fetal and adult brain, leukocytes
 CC and fetal fibroblasts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AB005217; BAA22433.1; -.
 CC FT NON TER 1 1
 CC FT NON TER 416 416
 CC SQ SEQUENCE 416 AA; 44829 MW; AA4B7F3A19522612 CRC64;
 Query Match 16.2%; Score 74.5; DB 1; Length 416;
 Best Local Similarity 31.5%; Pred. No. 7.1;
 Matches 28; Conservative 5; Mismatches 33; Indels 23; Gaps 4;
 QY 4 SKQPVSNVRAIQANINIPMGAFRPGAGOPRR-RKECTPEVERGCVPTSD-EKKKIP 48
 DB 192 SSGSLVYDQITFPONIPN-FTTAPGMPFRDWCPCPSQTGSRRAIESDSGEPP 250
 QY 49 TSDEE-----KKPIPGAKLPGPAVNL 70
 DB 251 TRDEHGGSGAGARGEIIPSPGSRVSLS 279
 RESULT 7
 BAT2 HUMAN
 ID BAT2 HUMAN STANDARD; PRT; 2142 AA.
 AC P48634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Large proline-rich protein BAT2 (HDA-B-associated transcript 2).
 GN BAT2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RX MEDLINE=90192810; PubMed=2156268;
 RA Banerji J., Sands J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 RT large proline-rich proteins with multiple repeated motifs and a
 RT single ubiquitin-like domain";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
 [2]
 RP SEQUENCE OF 1-1860 FROM N.A.
 RX MEDLINE=93272029; PubMed=8499947;
 RA Irit F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
 RA Cohen D.;
 DE "Dense Alu clustering and a potential new member of the NF kappa B

RT family within a 90 kilobase HLA class III segment.",
 CC Nat. Genet. 3:137-145(1993).
 CC -1- FUNCTION: UNKNOWN.
 CC -1- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).

DR EMBL; M33509; AAA35585.1; -;
 DR EMBL; M33518; AAA35586.1; -;
 DR EMBL; M33512; AAA35586.1; JOINED.
 DR EMBL; Z15025; CAA78744.1; -;
 DR PIR; B35098; B35098.
 DR PIR; S36152; S36152.
 DR Genbank; HGNC:13918; BAT2.
 DR MIM; 142580; -;
 KW Repeat.
 FT DOMAIN 519 524 POLY-PRO.
 FT 636 657 GLN-RICH.
 FT 684 688 POLY-PRO.
 FT 699 704 POLY-PRO.
 FT 814 821 POLY-PRO.
 FT 1340 1345 POLY-PRO.
 FT 1398 1403 POLY-GLY.
 FT 1436 1442 POLY-PRO.
 FT DOMAIN 1982 1991 POLY-PRO.
 FT 41 1795 4 X 57 AA TYPE A REPEATS.
 FT REPEAT 41 95 1-1.
 FT REPEAT 98 154 1-2.
 FT REPEAT 281 337 1-3.
 FT 1740 1795 1-4.
 FT 337 549 2 X TYPE B REPEATS.
 FT REPEAT 337 418 2-1.
 FT 476 549 2-2.
 FT 1899 2089 3 X 50 AA TYPE C REPEATS.
 FT REPEAT 1899 1948 3-1.
 FT 1965 2014 3-2.
 FT REPEAT 2040 2089 3-3.
 FT 57 57 R -> A (IN REF. 2).
 FT 109 109 Q -> S (IN REF. 2).
 FT 414 414 P -> PPHRGPAAGMGPP (IN REF. 2).
 FT 532 532 T -> K (IN REF. 2).
 FT 682 682 Q -> K (IN REF. 2).
 FT 730 730 E -> D (IN REF. 2).
 FT 750 750 L -> R (IN REF. 2).
 FT 834 834 A -> T (IN REF. 2).
 FT 1035 1035 G -> A (IN REF. 2).
 FT 1068 1068 M -> L (IN REF. 2).
 FT 1285 1285 P -> R (IN REF. 2).
 FT 1400 1400 G -> A (IN REF. 2).
 FT 1611 1611 T -> S (IN REF. 2).
 FT 1729 1729 G -> A (IN REF. 2).
 SQ SEQUENCE 2142 AA; 227840 MW; 32DDFL6B9B52420A CRC64;

Query Match 16.1%; Score 74; DB 1; Length 2142;
 Best Local Similarity 35.0%; Pred. No. 44;
 Matches 21; Conservative 8; Mismatches 15; Indels 16; Gaps 4;

QY 19 NIMPGAPRPAAGPAPRRKCTPVEEG-VPTSDDEKK-----PIGAKK-----LPGP 66
 DB 1147 SLREGALSPG-----PKRRRAPQVCPGWSPPAKSLAPKKPPTGTLPLPSKPEPLKEXLIPGP 1202

RESULT 8
 BTK HUMAN
 ID BTK_HUMAN
 AC 006187;
 DT 01-JUN-1994 (Rel. 29, Created)

STANDARD; PRT; 659 AA.

DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase BTK (EC 2.7.1.112) (Bruton's tyrosine kinase)
 DE (agammaglobulinemia tyrosine kinase) (ATK) (B cell progenitor
 DE kinase) (BPK).
 GN BTK OR ATK OR AGMX1 OR BPK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93140868; PubMed=8380905;
 RA Vetric D., Vorechovsky I., Sideras P., Holland J., Davies A.,
 RA Flinter F., Hammarstrom L., Kinnon C., Levinsky R.J., Bobrow M.,
 RA Smith C.I.E., Bentley D.R.;
 RT "The gene involved in X-linked agammaglobulinemia is a member of the
 RT src family of protein-tyrosine kinases.";
 RL Nature 361:226-233(1993).
 RN [2]
 RP ERRATUM.
 RA Vetric D., Vorechovsky I., Sideras P., Holland J., Davies A.,
 RA Flinter F., Hammarstrom L., Kinnon C., Levinsky R.J., Bobrow M.,
 RA Smith C.I.E., Bentley D.R.;
 RL Nature 364:362-362(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94377492; PubMed=8090769;
 RA Ohta Y., Haire R.N., Litman R.T., Fu S.W., Nelson R.P., Kretz J.,
 RA Kornfeld S.J., la Morena M., Good R.A., Litman G.W.;
 RT "Genomic organization and structure of Bruton agammaglobulinemia
 RT tyrosine kinase: localization of mutations associated with varied
 RT clinical presentations and course in X chromosome-linked
 RT agammaglobulinemia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9062-9066(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95012452; PubMed=7927535;
 RA Rohrer J., Parolini O., Belmont J.W., Conley M.E.;
 RT "The genomic structure of human BTK, the defective gene in X-linked
 RT agammaglobulinemia.";
 RL Immunogenetics 40:319-324(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95352959; PubMed=7626884;
 RA Oeltjen U.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
 RA Gibbs R.A.;
 RT "Sixty-nine kilobases of contiguous human genomic sequence containing
 RT the alpha-galactosidase A and Bruton's tyrosine kinase loci.";
 RL Mamm. Genome 6:334-338(1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX Oeltjen U.C., Malley T.M., Muzny D.M., Miller W., Gibbs R.A.,
 RA Belmont J.W.;
 RT Submitted (Nov-1996) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 1-442 FROM N.A.
 RX MEDLINE=93145329; PubMed=8425221;
 RA Tanaka S., Saitan D.C., Rawlings D.J., Parolini O., Allen R.C.,
 RA Kilsak I., Sparks R.S., Kubagawa H., Mohandas T., Quan S.,
 RA Belmont J.W., Cooper M.D., Conley M.E., Witte O.N.;
 RT "Deficient expression of a B cell cytoplasmic tyrosine kinase in
 RT human X-linked agammaglobulinemia.";
 RL Cell 72:279-290(1993).
 RN [8]
 RP PHOSPHORYLATION OF GTP21, AND MUTAGENESIS.
 RX MEDLINE=97165069; PubMed=9012831;
 RA Yang W., Desiderio S.;
 RT "BAF-135, a target for Bruton's tyrosine kinase in response to B cell
 RT receptor engagement.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:604-609(1997).
 RN [9]

RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-170.
 RX MEDLINE=97361821; PubMed=9218782;
 RA Hyvoenen M., Saraste M.;
 RT "Structure of the PH domain and Btk motif from Bruton's tyrosine
 RT kinase: molecular explanations for X-linked agammaglobulinemia.";
 RL EMBO J. 16:3396-3404(1997).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-170.
 RX MEDLINE=99216540; PubMed=10196129;
 RA Baraldi E., Carugo K.D., Hyvoenen M., Surdo P.L., Riley A.M.,
 RA Potter B.V.L., O'Brien R., Ladbury J.E., Saraste M.;
 RT "Structure of the PH domain from Bruton's tyrosine kinase in complex
 RT with inositol 1,3,4,5-tetrakisphosphate.";
 RL Structure 7:449-460(1999).
 RN [11]
 RP STRUCTURE BY NMR OF 209-275.
 RX MEDLINE=98153176; PubMed=9485443;
 RA Hansson H., Mattsson P.T., Allard P., Haapaniemi P., Vihinen M.,
 RA Smith C.I.E., Haerd T.;
 RT "Solution structure of the SH3 domain from Bruton's tyrosine kinase.";
 RL Biochemistry 37:2912-2924(1998).
 RN [12]
 RP REVIEW ON XLA VARIANTS.
 RX MEDLINE=96174621; PubMed=8594569;
 RA Vihinen M., Iwata T., Kinnon C., Kwan S.-P., Ochs H.D.,
 RA Vorechovsky I., Smith C.I.E.;
 RT "Btkbase, mutation database for X-linked agammaglobulinemia (XLA).";
 RL Nucleic Acids Res. 24:160-165(1996).
 RN [13]
 RP REVIEW ON XLA VARIANTS.
 RX MEDLINE=97169387; PubMed=9016530;
 RA Vihinen M., Belohradsky B.H., Haire R.N., Holinski-Feder E.,
 RA Kwan S.-P., Lappalainen I., Leheslahti H., Lester T., Meindl A.,
 RA Ochs H.D., Ollila J., Vorechovsky I., Weiss M., Smith C.I.E.;
 RT "Btkbase, mutation database for X-linked agammaglobulinemia (XLA).";
 RL Nucleic Acids Res. 25:166-171(1997).
 RN [14]
 RP VARIANTS XLA TRP-288; GLY-307; ASP-607 AND S-V-F-S-T-R-103 INS.
 RX MEDLINE=94214481; PubMed=8162056;
 RA Bradley L.A.D., Sweatman A.K., Lovering R.C., Jones A.M., Morgan G.,
 RA Levinsky R.J., Kinnon C.;
 RT "Mutation detection in the X-linked agammaglobulinemia gene, BTK,
 RT using single strand conformation polymorphism analysis.";
 RL Hum. Mol. Genet. 3:79-83(1994).
 RN [15]
 RP VARIANTS XLA HIS-28 AND TRP-288.
 RX MEDLINE=94214435; PubMed=8162018;
 RA de Weerts M., Menseink R.G.J., Kraakman M.E.M., Schuurman R.K.B.,
 RA Hendriks R.W.;
 RT "Mutation analysis of the Bruton's tyrosine kinase gene in X-linked
 RT agammaglobulinemia: identification of a mutation which affects the
 RT same codon as is altered in immunodeficient xid mice.";
 RL Hum. Mol. Genet. 3:161-166(1994).
 RN [16]
 RP SEQUENCE FROM N.A., AND VARIANTS XLA S-334; R-506; Q-520; W-562 AND
 RX MEDLINE=95152493; PubMed=7880320;
 RA Hagmann T.L., Chen Y., Rosen F.S., Kwan S.-P.;
 RT "Genomic organization of the Btk gene and exon scanning for mutations
 RT in patients with X-linked agammaglobulinemia.";
 RL Hum. Mol. Genet. 3:1743-1749(1994).
 RN [17]
 RP VARIANTS XLA D-113; C-361; Q-520; P-542; W-562; K-630 AND P-652.
 RX MEDLINE=95152494; PubMed=7849697;
 RA Conley M.E., Fitch-Hilgenberg M.E., Cleveland J.L., Parolini O.,
 RA Rohrer J.;
 RT "Screening of genomic DNA to identify mutations in the gene for
 RT Bruton's tyrosine kinase.";
 RL Hum. Mol. Genet. 3:1751-1756(1994).
 RN [18]
 RP VARIANTS XLA H-28; P-33; P-408; G-589; D-613 AND 260-Q-E-280 DEL.
 RX MEDLINE=95152522; PubMed=7849721;
 RA Zhu Q., Zhang M., Winkelstein J., Chen S.-H., Ochs H.D.;

RT "Unique mutations of Bruton's tyrosine kinase in fourteen unrelated
 RT X-linked agammaglobulinemia families.";
 RL Hum. Mol. Genet. 3:1899-1900(1994).
 RN [19]
 RP VARIANTS XLA E-430; Q-520; Q-525; P-562; V-582; G-589; E-594 AND
 RX MEDLINE=95108046; PubMed=7809124;
 RA Vihinen M., Vetric D., Maniar H.S., Ochs H.D., Zhu Q., Vorechovsky I.,
 RA Webster A.D.B., Notarangelo L.D., Nilsson L., Sowadski J.M.,
 RA Smith C.I.E.;
 RT "Structural basis for chromosome X-linked agammaglobulinemia: a
 RT tyrosine kinase disease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12803-12807(1994).
 RN [20]
 RP VARIANT XLA PHE-64, AND CHARACTERIZATION OF OTHER XLA VARIANTS.
 RX MEDLINE=95151728; PubMed=7849006;
 RA Vihinen M., Zvelebil J.J.M., Zhu Q., Broclmans R.A., Ochs H.D.,
 RA Zegers B.J.M., Nilsson L., Waterfield M.D., Smith C.I.E.;
 RT "Structural basis for pleckstrin homology domain mutations in
 RT X-linked agammaglobulinemia.";
 RL Biochemistry 34:1475-1481(1995).
 RN [21]
 RP VARIANTS XLA S-25; W-288; M-370; V-509; P-525; K-526; W-562
 RX AND R-594.
 RX MEDLINE=95227177; PubMed=7711734;
 RA Vorechovsky I., Vihinen M., de Saint Basile G., Honsova S.,
 RA Hammarstrom L., Mueller S., Nilsson L., Fischer A., Smith C.I.E.;
 RT "DNA-based mutation analysis of Bruton's tyrosine kinase gene in
 RT patients with X-linked agammaglobulinemia.";
 RL Hum. Mol. Genet. 4:51-58(1995).
 RN [22]
 RP VARIANTS XLA LYS-567; LEU-587 AND HIS-641.
 RX MEDLINE=95359977; PubMed=7633420;
 RA Jin H., Webster A.D.B., Vihinen M., Sideras P., Vorechovsky I.,
 RA Hammarstrom L., Bernatowska-Matuszkiewicz E., Smith C.I.E.,
 RA Bobrow M., Vetric D.;
 RT "Identification of Btk mutations in 20 unrelated patients with
 RT X-linked agammaglobulinemia (XLA).";
 RL Hum. Mol. Genet. 4:693-700(1995).
 RN [23]
 RP VARIANTS XLA PRO-33; GLN-520; CYS-641 AND GLY-302 DEL.
 RX MEDLINE=95359987; PubMed=7633429;
 RA Gaspar H.B., Bradley L.A.D., Katz F., Lovering R.C., Rolfman C.M.,
 RA Morgan G., Levinsky R.J., Kinnon C.;
 RT "Mutation analysis in Bruton's tyrosine kinase, the X-linked
 RT agammaglobulinemia gene, including identification of an insertional
 RT hotspot.";
 RL Hum. Mol. Genet. 4:755-757(1995).
 RN [24]
 RP VARIANTS XLA ASN-429 AND ARG-477.
 RX MEDLINE=96177680; PubMed=8634718;
 RN [25]
 RP Query Match 16.0%; Score 73.5; DB 1; Length 659;
 Best Local Similarity 32.4%; Pred. No. 14;
 Matches 22; Conservative 13; Mismatches 14; Indels 19; Gaps 5;
 Oy 23 GARPPAGQPPRRKCTPPYBEGVPTSDDE---KKPIPGAKLLPGPA---VNLSEIOMI 76
 Db 173 GILRPSS---HRRTKRP-----LPPPEBDQILKKPLP-----PAPAAPVTSLSKV 219
 Oy 77 KSELKYVP 84
 Db 220 VALDYMP 227
 RESULT 9
 HXAA HUMAN STANDARD; PRT; 393 AA.
 ID HXAA HUMAN
 AC P31260; O15949; O43605; O43370;
 DT 01-JUN-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Homeobox protein Hox-A10 (Hox-1H) (Hox-1.8) (PL).

GN HOXA10 OR HOX11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RX MEDLINE=91288229; PubMed=1676505;
 RA Lowrey P., Corral J.C., Detmer K., Lebeau M.M., Deaven L.,
 RT Lawrence H.J., Largman C.;
 "A human Hox 1 homeobox gene exhibits myeloid-specific expression of
 alternative transcripts in human hematopoietic cells.";
 RL Nucleic Acids Res. 19:3443-3449(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ml X., Winters J.L., Stevens D.B., Fleischman R.A.;
 RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Bradshaw H., Hinde K., Keppeler D.;
 RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 301-383 FROM N.A.
 RX MEDLINE=90046832; PubMed=2573064;
 RA Shen W.-F., Largman C., Lowrey P., Corral J.C., Detmer K.,
 RT Hauser C.A., Simonitch T.A., Hack F.M., Lawrence H.J.;
 " lineage-restricted expression of homeobox-containing genes in human
 hematopoietic cell lines ";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8536-8540(1989).
 RN [5]
 RP SEQUENCE OF 319-384 FROM N.A.
 RX MEDLINE=90098976; PubMed=2574852;
 RA Acampora D., D'Esposito M., Faiella A., Pannese M., Migliaccio E.,
 RT Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
 "The human HOX gene family.";
 RL Nucleic Acids Res. 17:10385-10402(1989).
 RN [6]
 RP SEQUENCE OF 332-370 FROM N.A.
 RX MEDLINE=94145486; PubMed=7906121;
 RA Castronovo V., Kusaka M., Charlet A., Gielen J., Sobel M.;
 RT "Homeobox genes: potential candidates for the transcriptional control
 of the transformed and invasive phenotype.";
 RL Biochem. Pharmacol. 47:137-143(1994).
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
 SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
 CC BINDS TO THE DNA SEQUENCE 5'-AAATATTTTATTTAC-3'.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, 1/PU1 (SHOWN HERE) AND 2/PU2;
 ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A NUMBER OF
 FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58430; CAB86198.1; ALT_FRAME.
 DR EMBL; AF040714; AAB96917.1; --
 DR EMBL; AC004080; -- NOT ANNOTATED_CDS.
 DR EMBL; M30599; AAA36006.1; --
 DR EMBL; S69027; AAD14030.1; --
 DR EMBL; S69029; AAD14031.1; --
 DR PIR; S14930; S14930.
 DR PIR; S26402; S26402.
 DR PIR; A34425; A34425.
 DR HSSP; P02833; GANT.
 DR TRANSFAC; T01713; --.

DR Genew; HGNC:5100; HOXA10.
 DR MIM; 142957; --
 DR InterPro; IPR001356; Homeobox.
 DR Pfam; PF00046; homeobox; 1.
 DR PRINTS; PR00024; HOMEBOX.
 DR ProDom; PD000010; Homeobox; 1.
 DR SMART; SM00389; HOX; 1.
 DR PROSITE; PS00027; HOMEBOX_1; 1.
 DR PROSITE; PSS0071; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
 KW Transcription regulation; Alternative splicing.
 FT DOMAIN 26 39
 FT DOMAIN 78 89
 FT DOMAIN 115 120
 FT DOMAIN 124 133
 FT DOMAIN 208 215
 FT DOMAIN 256 259
 FT DNA_BIND 319 378
 FT VARSPPLIC 1 299
 FT VARSPPLIC 300 302
 FT CONFLICT 1 6
 FT CONFLICT 69 69
 FT CONFLICT 136 137
 FT CONFLICT 332 332
 FT CONFLICT 334 334
 FT CONFLICT 370 370
 SQ SEQUENCE 393 AA; 40536 MW; 20F89542582D6F25 CRC64;
 Query Match 15.1%; Score 69.5; DB 1; Length 393;
 Best Local Similarity 30.4%; Pred. No. 19;
 Matches 21; Conservative 6; Mismatches 33; Indels 9; Gaps 2;
 QY 23 GARRPQA---GQPP-----RRKCTPEVEGVPTSDSEKKIPGAKKLGPAVNLSEI 73
 DB 84 GGLGPGAGHYGSPIDILMDLPDRSCRMPPDPGPPPPQPPPPQPPAPQATSCSFA 143
 QY 74 QNKKSELYK 82
 DB 144 QNKKSESY 152
 RESULT 10
 SE5 RAT STANDARD; PRT; 825 AA.
 AC Q63003;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SE5 antigen.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brain;
 RX MEDLINE=96015159; PubMed=8537300;
 RA Suzuki E., Kojima N., Yoshimura K., Uyemura K., Obata K., Akagawa K.;
 RT "Cloning and sequence analysis of cDNA for a possible DNA-binding
 protein SE5 in the nervous system.";
 RL J. Biochem. 118:122-128(1995).
 CC -1- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: D37934; BAA07153.1; -.
KW DNA-binding; Nuclear protein; Antigen.
SQ SEQUENCE 825 AA; 86831 MW; AF67FE2FD55BDF CRC64;

Query Match
Best Local Similarity 15.1%; Score 69.5; DB 1; Length 825;
Matches 23; Conservative 6; Mismatches 19; Indels 25; Gaps 3;

QY 21 PMGAFRPGAGPPRRKCTPEVEGV-----PTSDEKKPIPKAKLP 65
DB 139 PEGAF-PCGG-----PAEEGVPRPRAPPPDPGAPRPPDPGLPLPGSGEKPT 188
QY 66 PAVNLSEIIONIKS 78
DB 189 FVYQVSTEOILMS 201

RESULT 11
MSX3_MOUSE STANDARD; PRT; 204 AA.
ID MSX3_MOUSE
AC P70354; P70246;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein MSX-3.
GN MSX3 OR MSX-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97042057; PubMed=8887328;
RA Wang W., Chen X., Xu H., Luftin T.;
RT "Mx3: a novel murine homologue of the Drosophila msh homeobox gene
RT restricted to the dorsal embryonic central nervous system.";
RL Mech. Dev. 58:203-215(1996).

RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97014264; PubMed=8861099;
RA Shimeid S.M., McKay I.J., Sharpe P.T.;
RT "The murine homeobox gene Mx3 shows highly restricted expression in
RT the developing neural tube.";
RL Mech. Dev. 55:201-210(1996).

RN [3]
RP SEQUENCE OF 73-204 FROM N.A.
RX MEDLINE=97014264; PubMed=8861099;
RA Shimeid S.M., McKay I.J., Sharpe P.T.;
RT "The murine homeobox gene Mx3 shows highly restricted expression in
RT the developing neural tube.";
RL Mech. Dev. 55:201-210(1996).

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: RESTRICTED TO THE DORSAL EMBRYONIC CENTRAL
CC NERVOUS SYSTEM.
CC -1- SIMILARITY: BELONGS TO THE MSH HOMEOBOX FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U62523; AAB49935.1; -.
DR EMBL: AF060229; AAC15459.1; -.
DR EMBL: X96518; CAA65367.1; -.
DR HSSP: P14653; 1B72.
DR MGD; MGI:106587; Mx3.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.

```

```

DR PRINTS; PR00024; HOMEOBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS00071; HOMEOBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Multigene family.
FT DNA BIND 87
FT CONFLICT 75 146 HOMEOBOX
FT CONFLICT 75 75 S -> I (TN REF. 3).
SQ SEQUENCE 204 AA; 21971 MW; BC032991DCCB4AA1 CRC64;

Query Match
Best Local Similarity 15.0%; Score 69; DB 1; Length 204;
Matches 20; Conservative 12; Mismatches 37; Indels 8; Gaps 2;

QY 15 QANINIPMGAFRPGAGPPRRKCTPEVEGVPTSDEK---KPIPKAKLP 71
DB 46 ELGVERPLGASKPGAMPVPAHSCPPRAPSPPCTLRKHTNKKP-----KTFPTAQL 100
QY 72 EIQNIKSELKVPKAEQ 88
DB 101 ALERKPHQYLSIAER 117

RESULT 12
BCA1_MOUSE STANDARD; PRT; 874 AA.
ID BCA1_MOUSE
AC Q61140; Q60869;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE CRK-associated substrate (P130CAS) (Breast cancer anti-estrogen
DE resistance 1 protein).
GN BCAR1 OR CRKAS OR CAS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS CAS-A AND CAS-B), AND INTERACTION WITH
RP FOCAL ADHESION KINASE.
RC TISSUE=Embryo;
RX MEDLINE=96068679; PubMed=7479864;
RA Polte T.R., Hanks S.K.;
RT "Interaction between focal adhesion kinase and Crk-associated tyrosine
RT kinase substrate p130Cas.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:10678-10682(1995).

RN [2]
RP INTERACTION WITH NEPHROCYSTIN.
RX MEDLINE=20249316; PubMed=10739664;
RA Donaldson V.C., Dempsey P.J., Reddy S., Bouton A.H., Coffey R.J.;
RA Hanks S.K.;
RT "Crk-associated substrate p130(Cas) interacts with nephrocystin and
RT both proteins localize to cell-cell contacts of polarized epithelial
RT cells.";
RL Exp. Cell Res. 256:168-178(2000).

CC -1- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE
CC FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION
CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY). HAS
CC BEEN SHOWN TO BE ESSENTIAL IN CARDIOVASCULAR DEVELOPMENT DURING
CC EMBRYOGENESIS.
CC -1- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1,
CC ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL
CC (BY SIMILARITY). INTERACTS WITH NEPHROCYSTIN AND PIK2B.
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS. UNPHOSPHORYLATED FORM
CC LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO THE MEMBRANE UPON
CC TYROSINE PHOSPHORYLATION (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CAS-B (SHOWN HERE) AND CAS-A;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTATE DOMAIN) CONTAINING
CC MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN
CC CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-
CC BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE

```

HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOHYPHAL GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.

-1- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM RESPONSE ELEMENT (SRE) (BY SIMILARITY).

-1- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH REGION OF FOCAL ADHESION KINASE 1.

-1- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE YDYVHL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN-MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX (BY SIMILARITY).

-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

-1- SIMILARITY: BELONGS TO THE CAS FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U08653; AAA93381.1; -
EMBL; U28151; AAA93248.1; -
HSSP; P07751; 1BK2.
MGD; MGI:108091; CrkAa.
InterPro; IPR001452; SH3.
Pfam; PF00018; SH3; 1.
PRINTS; PR00452; SH3DOMAIN.
ProDom; PD000066; SH3; 1.
SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR Phosphorylation; SH3 domain; SH3-binding; Cell adhesion; Alternative splicing
KM DOMAIN 6 26 SH3.
FT DOMAIN 74 87 PRO-RICH.
FT DOMAIN 119 420 SUBSTRATE FOR KINASES (BY SIMILARITY).
FT SITE 426 618 SER-RICH.
FT SITE 639 647 SH3-BINDING (POTENTIAL).
FT DOMAIN 750 800 DIVERGENT HELIX-LOOP-HELIX MOTIF.
FT VARSPLIC 1 4 MTP -> MKTL (IN ISOFORM CAS-A).
SQ SEQUENCE 874 AA; 94256 MW; 5B9EDD653BDBA CRC64;

Query Match 15.0%; Score 69; DB 1; Length 874;
Best local Similarity 29.9%; Pred. No. 49;
Matches 23; Conservative 10; Mismatches 30; Indels 14; Gaps 4;

QY 19 NIPWGAIRPGAG---QPPRRKCTPEVEEG-----VPTSDERKQIPGAKLPGP 66
DB 377 DVPPGLRRPFGTLVYDVRERVLPEEVADGSVDVGVAVPPAERE-APTDG-KRLSAS 434
QY 67 ANVLSEIQLKSEIKYV 83
DB 435 STGSTRSSGASASLEVV 451

RESULT 13
BCAL RAT STANDARD; PRT; 968 AA.
ID Q63767; Q63766; 30-MAY-2000 (Rel. 39, Created)
AC Q63767; Q63766; 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE CRK-associated substrate (P130CAS) (Breast cancer anti-estrogen resistance 1 protein).
GN BCAR1 OR CRKAS OR CAS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=94349922; PubMed=8070403;
RA Sakai R., Imanishi A., Hirano N., Ogawa S., Tanaka T., Mano H.,
RA Yezaki Y., Hirai H.,
RT "A novel signaling molecule, p130, forms stable complexes in vivo with v-Crk and v-Src in a tyrosine phosphorylation-dependent manner.";
RL EMBO J. 13:3748-3756 (1994).
RN [2]
RP TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.
RX MEDLINE=98005058; PubMed=9350983;
RA Tachibana K., Urano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,
RA Hirai H., Morimoto C.;
RT "Tyrosine phosphorylation of Crk-associated substrates by focal adhesion kinase. A putative mechanism for the integrin-mediated tyrosine phosphorylation of Crk-associated substrates.";
RL J. Biol. Chem. 272:29083-29090 (1997).
CC -1- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.
CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).
CC -1- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1, ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL. INTERACTS WITH NEPHROCYSTIN AND PTK2B (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS. UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE) AND A SHORT ISOFORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG, INTESTINE AND TESTIS.
CC -1- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-BINDING SITES PUTATIVELY BIND CRK, NCK AND ABL SH2 DOMAINS. THE HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOHYPHAL GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.
CC -1- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM RESPONSE ELEMENT (SRE).
CC -1- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH REGION OF FOCAL ADHESION KINASE 1.
CC -1- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE YDYVHL MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN-MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.
CC -1- DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF SOME CELL TYPES.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE CAS FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D29766; BAA06169.1; -
EMBL; D29766; BAA06170.1; -
HSSP; P07751; 1BK2.
InterPro; IPR001452; SH3.
Pfam; PF00018; SH3; 1.
PRINTS; PR00452; SH3DOMAIN.
ProDom; PD000066; SH3; 1.
SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR Phosphorylation; SH3 domain; SH3-binding; Cell adhesion; Alternative splicing
KM DOMAIN 97 159 SH3.
FT DOMAIN 168 181 PRO-RICH.
FT DOMAIN 213 514 SUBSTRATE FOR KINASES.

```

FT DOMAIN 520 712 SER-RICH.
FT SITE 733 741 SH3-BINDING (POTENTIAL).
FT DOMAIN 844 894 DIVERGENT HELIX-LOOP-HELIX MOTIF.
FT VARSPLIC 5 98 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 968 AA; 104262 MW; E861641BFP68D377 CRC64;

Query Match 15.0%; Score 69; DB 1; Length 968;
Best Local Similarity 29.9%; Pred. No. 55;
Matches 23; Conservative 10; Mismatches 30; Indels 14; Gaps 4;

QY 19 NIPMGAFRPGAG--OPRRKCTPEVEEG-----VPTSDDEKKPIPGAKLPGP 66
DB 471 DVEGLARRPGGTGYDPRREVLPPEVADSGVIDDGYAVPAPARR-APTDG-KRLSAS 528
QY 67 AVNLSEIOTIKSELKYY 83
DB 529 STGSTRSSQSASLLEV 545

RESULT 14
CES2 HUMAN STANDARD; PRT; 1484 AA.
ID CES2_HUMAN 09BXC3; 09COC3; 096P58;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cat eye syndrome critical region protein 2.
GN CECR2 OR KIA11740.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA MEDLINE=21275466; PubMed=11381032;
RA Footz T.K., Brinkman-Mills P., Banting G.S., Maier S.A., Riazzi M.A.,
RA Bridgland L.J., Hu S., Biren B., Minoshima S., Shimizu N., Pan H.,
RA Nguyen T., Fang F., Fu Y., Ray L., Wu H., Shail S., Phan S., Yao Z.,
RA Chen F., Huan A., Hu P., Wang Q., Loh P., Qi S., Roe B.A.,
RA McDermid H.E.;
RT "Analysis of the cat eye syndrome critical region in humans and the
RT region of conserved synteny in mice: a search for candidate genes at
RT or near the human chromosome 22 pericentromere.";
RL Genome Res. 11:1053-1070(2001).
RN [2]
RA MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
RN [3]
RA MEDLINE=21686162; PubMed=11827465;
RA Liu L., McKeehan W.L.;
RT "Sequence analysis of LRPPRC and its SEC1 domain interaction partners
RT suggests roles in cytoskeletal organization, vesicular trafficking,
RT nucleocyosolic shuttling, and chromosome activity.";
RL Genomics 79:124-136(2002).
RN [4]
RA MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Mose H.W.M., McNab D., Frame M.C.;
RT "TSSB SPECIFICITY. Highly expressed in skeletal muscle, thymus,
RT placenta and lung. Expressed at lower level in brain, heart,
RT colon, spleen, kidney."

```

```

CC -1- MISCELLANEOUS: Candidate gene for the Cat Eye Syndrome (CES), a
CC developmental disorder associated with the duplication of a 2 Mb
CC region of 22q11.2. Duplication usually takes in the form of a
CC supernumerary bisatellited isodicentric chromosome, resulting in
CC four copies of the region (represents an inv dup(22)(q11)). CES is
CC characterized clinically by the combination of coloboma of the
CC iris and anal atresia with fistula, downslanting palpebral
CC fissures, preauricular tags and/or pits, frequent occurrence of
CC heart and renal malformations, and normal or near-normal mental
CC development.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF336133; AAK15343.1; -.
CC DR EMBL; AB051527; BAB21831.1; -.
CC DR EMBL; AF411609; AAL07393.1; -.
CC DR HSSP; 092831; 1B91.
CC DR Genew; HGNC:1840; CECR2.
CC DR InterPro; IPR001487; Bromodomain.
CC DR Pfam; PF00439; bromodomain; 1.
CC DR PRINTS; PR00503; BROMODOMAIN.
CC DR SMART; SM00297; BROMO; 1.
CC DR PROSITE; PS00633; BROMODOMAIN_1; 1.
CC DR PROSITE; PSS0014; BROMODOMAIN_2; 1.
CC KM Alternative splicing.
CC FT DOMAIN 451 521 BROMODOMAIN.
CC FT FT 333 337 POLY-GLU.
CC FT FT 611 614 POLY-SER.
CC FT DOMAIN 1250 1253 POLY-PRO.
CC FT VARSPLIC 291 318 MISSING (IN ISOFORM B).
CC FT VARSPLIC 519 526 EYKMSDN -> GKGRSLC (IN ISOFORM B).
CC FT VARSPLIC 527 1484 MISSING (IN ISOFORM B).
CC FT CONFLICT 370 389 MISSING (IN REF. 2).
CC FT CONFLICT 1029 1029 C -> S (IN REF. 2).
CC FT CONFLICT 1045 1045 R -> W (IN REF. 2).
CC SQ SEQUENCE 1484 AA; 164214 MW; 049AAB44E51AF63F CRC64;

Query Match 15.0%; Score 69; DB 1; Length 1484;
Best Local Similarity 32.7%; Pred. No. 85;
Matches 18; Conservative 11; Mismatches 20; Indels 6; Gaps 3;

QY 18 INIMGAFRPGAGOPRRKCTPEVEGVPTSDDEKKPIPGAKLPGP VNVNSE 72
DB 943 MSVTVASPKPALGNPGR---APENSEAQEPEND-QAEPLGAEKP-PGVGTSE 991

RESULT 15
VGLT_HSV2H STANDARD; PRT; 372 AA.
ID VGLT_HSV2H
AC P13291;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Glycoprotein I.
GN GI OR U57.
OS Herpes simplex virus (type 2 / strain HG52).
OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10315;
RN [1]
RA MEDLINE=87111457; PubMed=3027242;
RA McGeoch D.J., Mose H.W.M., McNab D., Frame M.C.;
RT "TSSB SPECIFICITY. Highly expressed in skeletal muscle, thymus,
RT placenta and lung. Expressed at lower level in brain, heart,
RT colon, spleen, kidney."
RT Identification of the gene encoding glycoprotein G, and evolutionary

```

```

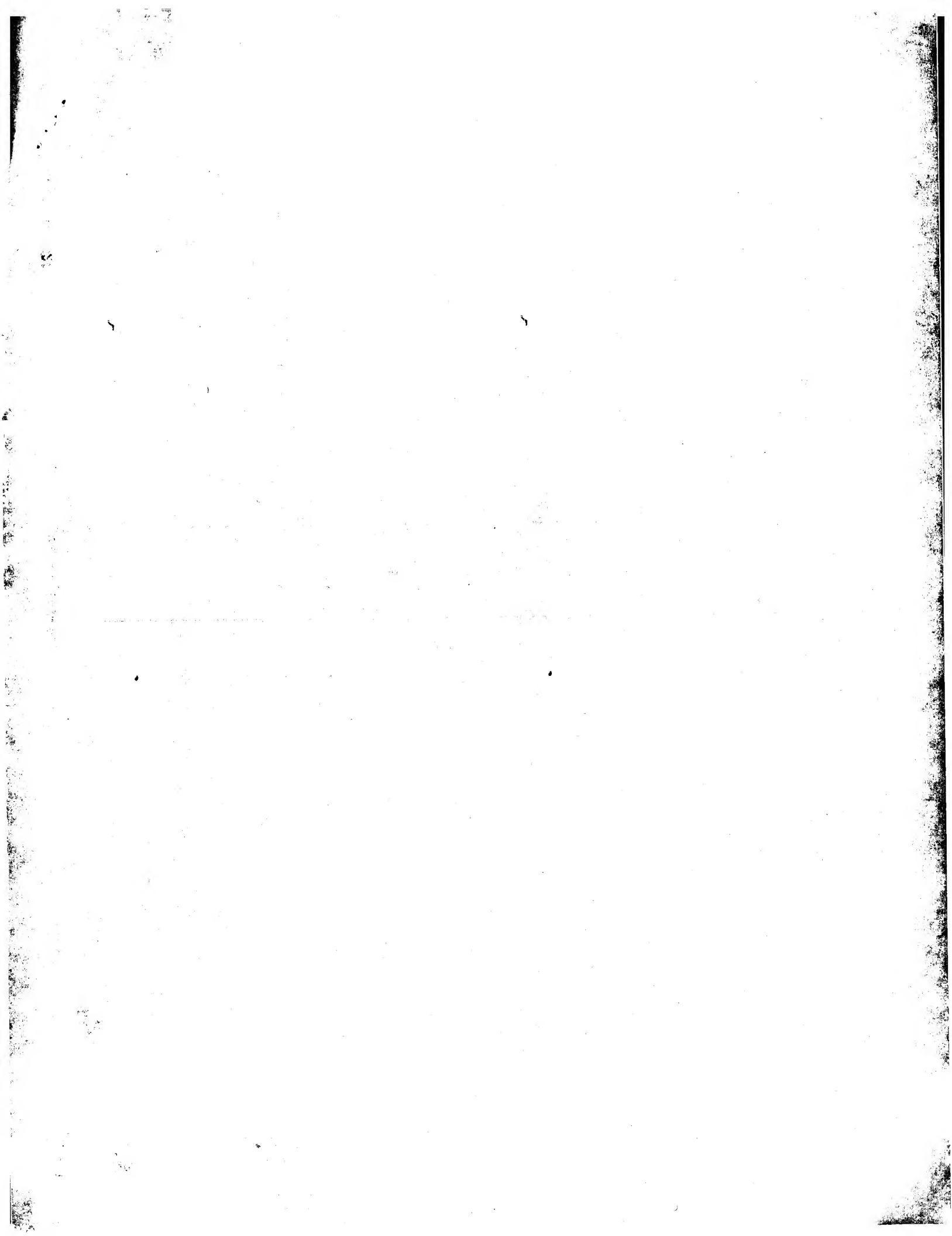
RT comparisons."
RL J. Gen. Virol. 68:19-38(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEF-1997) to the EMBL/GenBank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC 2: GH, GB, GC, GD, GI, AND GE.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS I, TO VZV GP1V,
CC AND TO PRV GP63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X04798; CA28485.1; -.
DR EMBL, Z86099; CAB06714.1; -.
DR PIR, F43674; F43674.
DR InterPro; IPR002874; Herpes_gf.
DR Pfam; PF01688; Herpes_gf; 1.
KV Glycoprotein.
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 372 AA; 39558 MM; COD4A22CAB16E8D7 CRC64;

Query Match 14.9%; Score 68.5; DB 1; Length 372;
Best Local Similarity 31.9%; Pred. No. 22;
Matches 15; Conservative 5; Mismatches 20; Indels 7; Gaps 1;

Qy 21 PMGAFRPGAGOPPRRREKTPPEVSEGVPTSDSEKKRIPGAKKLPGPA 67
Db 196 PSSVYTPGASRPTPPRTT-----PPSSPRDPTPAPGDTGTAPAPA 235

```

Search completed: April 9, 2003, 12:36:14
 Job time : 15.0645 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:33:21 Search time 22.3548 Seconds
(without alignments)
378.434 Million cell updates/sec

Title: US-09-647-019-4

Perfect score: 459

Sequence: 1 MNMSKQPVSNVRAIQANINI.....NLSEIONIKSELYPKAEQ 88

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: PIR 73:.*
2: p1r1:.*
3: p1r2:.*
4: p1r3:.*
5: p1r4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.5	18.6	1209	T00373	hypothetical prote
2	79	17.2	1239	VHWEV	structural polypro
3	79	17.2	1241	S26373	genome polypeptide
4	78.5	17.1	7962	I38346	elastic titin - hu
5	78	17.0	1242	S72350	structural polypro
6	77.5	16.9	1240	VHWEV	structural polypro
7	77.5	16.9	1240	A56605	structural polypro
8	77	16.8	458	T04106	B-type cyclin homo
9	77	16.8	479	T04104	B-type cyclin homo
10	76.5	16.7	990	T14756	hypothetical prote
11	76	16.6	294	A12016	hypothetical prote
12	75.5	16.6	3534	T42567	legumen protein 2
13	75.5	16.4	1072	A37127	microtubule-associ
14	74.5	16.2	1171	T13065	PIR82 protein - fi
15	74	16.1	1870	S37671	MHC class III hist
16	74	16.1	1872	S37652	MHC class III hist
17	74	16.1	2142	B35098	MHC class III hist
18	73.5	16.0	659	A45184	protein-tyrosine k
19	72	15.7	222	G86168	hypothetical prote
20	71.5	15.6	307	G87378	hypothetical prote
21	71.5	15.6	307	A82596	comb protein (limp
22	71.5	15.6	961	G90053	hypothetical prote
23	70.5	15.4	950	T30692	probable mRNA guan
24	70	15.3	151	D84688	hypothetical prote
25	69.5	15.3	270	H83619	hypothetical prote
26	69.5	15.1	384	T17229	hypothetical prote
27	69.5	15.1	460	JE0367	zinc finger protei
28	69.5	15.1	598	A75531	hypothetical prote
29	69.5	15.1	825	JC4163	DNA-binding protei

30	69	15.0	408	2	T32767	hypothetical prote
31	69	15.0	968	2	S46992	protein p130 - rat
32	68.5	14.9	372	2	F43574	US7 protein - huma
33	68.5	14.9	880	2	B87222	probable iron-sulf
34	68	14.8	170	2	T48144	hypothetical prote
35	68	14.8	249	2	JH0629	cleavage signal-1
36	68	14.8	514	2	I52703	42K membrane glyco
37	68	14.8	514	2	A33879	antipeptidease y8c
38	68	14.8	614	2	T16268	hypothetical prote
39	68	14.8	710	2	D96728	hypothetical prote
40	68	14.8	745	2	T12528	hypothetical prote
41	68	14.8	1415	2	T21244	zyg-9 protein - Ca
42	67.5	14.7	422	2	T24865	hypothetical prote
43	67.5	14.7	854	2	S02003	neurofilament trip
44	67.5	14.7	2440	2	S39162	transcription coac
45	67.5	14.7	2441	2	S39161	CREB-binding prote

ALIGNMENTS

RESULT 1

T00373 hypothetical protein KIA0649 - human

C/Species: Homo sapiens (man)

C/Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C/Accession: T00373

R/Hit: Kawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N

DNA Res. 5, 169-176, 1998

A/Title: Prediction of the coding sequences of unidentified human genes. X. The complet

A/Reference number: Z14142; MUID:98403880; PMID:9734811

A/Accession: T00373

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-1209 <ISH>

A/Cross-references: EMBL:AB014549; NID:G3327111; PIDN:BAA31624.1; PID:G3327112

A/Experimental source: brain

C/Genetics:

A/Note: KIA0649

Query Match	Best Local Similarity	Score	DB 2;	Length	1209;
Matches 27; Conservative	35.1%;	Pred. No. 2.6;	6; Mismatches 33; Indels 11; Gaps 2;		
QY 13 AIDANINIPMGAPRPAAG-----QPPR---KECTBEVEGVPTSDERKPIPGAK 61					
DB 136 AIDBYIKAKSGAQPAGAGPAGAPSPRACGSRCKEPAPASAPALCPKLVPGSG 195					
QY 62 KLRGPAVNLSEIONIKS 78					
DB 196 GGRGQVGSKDGAS 212					

RESULT 2

VHWEV

structural polyprotein - eastern equine encephalomyelitis virus (strain 82V-2137)

N/contains: 6K protein; coat protein C; membrane glycoprotein E1; membrane glycoprotein

C/Species: eastern equine encephalomyelitis virus

C/Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999

C/Accession: A26816

R/Chang, G.J.J.; Trent, D.W.

J. Gen. Virol. 68, 2129-2142, 1987

A/Title: Nucleotide sequence of the genome region encoding the 26S mRNA of eastern equi

A/Reference number: A26816; MUID:87282265; PMID:2886548

A/Accession: A26816

A/Molecule type: mRNA

A/Cross-references: EMBL:X05816; NID:G62074; PIDN:CAA29261.1; PID:G62075

C/Superfamily: togavirus structural polyprotein

C/Keywords: coat protein; glycoprotein; transmembrane protein

F/1-259/Product: coat protein C #status predicted <CPC>

F/260-322/Product: membrane glycoprotein E1 #status predicted <MG3>

F/261-277/Domain: transmembrane #status predicted <TM1>


```
Query Match      16.9%; Score 77.5; DB 1; Length 1240;
Beet Local Similarity 27.4%; Pred. No. 16;
Matchs 17; Conservative 14; Mismatches 24; Indels 7; Gaps 1;
```

```

RESULT 7
A:Accession: A56605
C:structural polypeptide - eastern equine encephalomyelitis virus (strain 4789)
C:Species: eastern equine encephalomyelitis virus
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 26-Aug-1999
C:Accession: A56609
R:Weaver, S.C.; Hagenbaugh, A.; Bellow, L.A.; Calisher, C.H.
Arch. Virol. 127, 305-314, 1992
A:Title: Genetic characterization of an antigenic subtype of eastern equine encephalomyelitis virus
A:Reference number: A56605; MUID:93090093; PMID:1280945
A:Accession: A56605
A:Status: Preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1242 <WEA>
A:Cross-references: GB:I20951; NID:G405814; PIDN:AAA02897.1; PID:G305047
A:Note: sequence inconsistent with nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:119924, NCBIPI:119931)
C:Superfamily: togavirus structural polypeptide
C:Keywords: polypeptide

```

	Query Match	16.9%	Score 77.5	DB 2,	Length 1242;
	Best Local Similarity	27.4%	Pred. No. 16;		
	Matches 17,	Conservative 14,	Mismatches 24;	Indels 7;	Gaps 1;
Oy	8	USNYAIAANINPGACGAGCGPPRKECTPEVEBG-----VPPSPDEKKPTGCA	60		
		::::: :			
Db	41	IEDURSRANLTIKORAPPPGPAPKAKPAVKRPKPDAOKKKRPPPPPKKKKKPCKGK	100		
		::: :			

QY 61 KK 62
Db 101 RQ 102

RESULT 8
T04106 B-type cyclin homolog (clone Cyc2me3) - maize
C|Species: Zea mays (maize)
C|Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C|Accession: T04106
R|Sun, Y.; Flannigan, B.A.; Madison, J.T.; Setter, T.L.
Gene 195, 167-175, 1997
A|Title: Alternative splicing of cyclin transcripts in maize endosperm.
A|Reference number: 215213; MUID:97449292; PMID:9305761
A|Accession: T04106
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: mRNA
A|Residues: 1-458 <SUN>
A|Cross-references: EMBL:U66662, NID:G1546054, PIDD:AA872019.1; PTD:G1546055
A|Experimental source: cultivar Pioneer 3925
C|Superfamily: cyclin
C|Keywords: alternative splicing

Query Match 16.8% Score 77; DB 2; Length 458;
 Best Local Similarity 35.7%; Pred. No. 6.2;
 Matches 20; Conservative 8; Mismatches 26; Indels 2; Gaps 1.

QY 10 NVRAIQANINIFMGAFRCAGQPPRRKCC-TPEVEGVPTSDDEKDPDPAKQL 63
 DB 124 NDRAAPARQKFLDVFNNRNGKAVLKKECKVAPEYVLSIPDSEKKEKKSFRPGQGY 179

RESULT 9
T04104
B-type cyclin homolog - maize
C.Species: Zea mays (maize)
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C.Accession: T04104
R.Sun, Y.; Flannigan, B.A.; Madison, J.T.; Setter, T.L.
Gene 195, 167-175, 1997
A.Title: Alternative splicing of cyclin transcripts in maize endosperm.
A.Reference number: Z15213, MUID:97444292, PMID:9305761
A.Accession: T04104
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-479 <SD>
A.Cross-references: EMBL:U66607, NID:G1545870, PIDN:ABE7020.1, PID:G1545871
A.Experimental source: cultivar Pioneer 3925
C.Genetics:
A.Introns: 195/3; 434/2
C.Superfamily: cyclin

Query Match 16.8%; Score 77; DB 2; Length 479;
 Best Local Similarity 35.7%; Pred. No. 6.5;
 Matches 20; Conservative 8; Mismatches 26; Indels 2; Gaps 1.

Cy 10 NVRLQAININIPMGAFRPGAGOPPRKRC--TPVEBGGVPTSDDEKKIPGAKTL 63
 Db 124 NDRAAPNARQKELDVFNRNGAVLTKCKCKAPEIVISIPSEKKKSPFGSGCY 179

Qy 10 NVRAIQANINIPMGAFRPGAGQPPRRKEC--TTEVBEGVPPPTSDSEKKPIPGAKUL 63
| | | : : : | | | : : : | : :
Db 124 NDRAAPANQKPLDAVFNRNGKAVLKECKVKKEVIIVSIPOSEKEKSKKEPGGQKV 179

RESULT 10
T14756

hypothetical protein DKFZ564F0923.1 - human (fragment)
C|Species: Homo sapiens (man)
C|Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C|Accession: T14756
R|Wambutt, R.; Haubner, D.; Mewes, H.W.; Gassenhuber, U.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A|Reference number: Z18181

```

1 A:Status: preliminary
2 A:Molecule type: mRNA
3 A:Residues: 1-990 <PAM>
4 A:Cross-references: EMBL:AL110210
5 A:Experimental source: fetal brain, clone DKFZ564F0923.1
6 C:Genetics:
7 C1:Note: DKFZ564F0923.1

```

```

Query Match          16.7%; Score 76.5; DB 2; Length 990;
Best Local Similarity 27.0%; Pred. No. 16;
Matches 24; Conservative 13; Mismatches 23; Indels 29; Gaps 4;

Oy 7 PVSNNRAIQANINI-PMG-----AFRPAQOPPRKRCETBEVGVAPDSEKKPIRG 59
   ||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 848 PISIQATITACISINPFGLESPVASLGPAPP-----GLPPASLPSTPIPS 896
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 60 AKKLGPANLSEIQNIKSELKYVPKAEQ 88
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 897 SSPP-----LSSPLPEADPKE 914
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Qy      60 AKKLPGAVNLSEIQNIKSELKTVPKAEQ 88
          : | | | | : :
Db      897 SSPP-----LSSPLPEAPQPK 914

```

RESULT 11

hypothetical protein alr1687 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A12016
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuch
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata,
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An.
A:Reference number: AB1807, MUID:21595285, PMID:11759840

A/Accession: A12016
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-294 <NR>
 A/Cross-references: GB:BA000019; PIDN:BA878053.1; PID:g17135507; GSPDB:GN00179
 A/Experimental source: strain PCC 7120
 C/Genetics:
 A/Genes: alr1687

Query Match 16.6%; Score 76; DB 2; Length 294;
 Best Local Similarity 31.4%; Pred. No. 4.8;
 Matches 22; Conservative 7; Mismatches 37; Indels 4; Gaps 2;

QY 7 PVSNNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPPTSDDEKKIPGAKKLPGP 66
 DB 132 PVEVPPTPTPTAP--EPSPLEPPPTATPTPEPTQIQP-DEPEPTPEPATPTPTP 187
 QY 67 AVNLSEIQNI 76
 DB 188 PIQLTPEQTL 197

RESULT 12
 T42567
 tegument protein 24 - equine herpesvirus 4 (strain NS80567)
 C/Species: equine herpesvirus 4
 A/Variety: strain NS80567
 C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T42567
 R/RefSeq: E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 J. Gen. Virol. 79, 1197-1203, 1998

A/Title: The DNA sequence of equine herpesvirus-4.
 A/Reference number: 222173; MUID:98264497; PMID:9603335
 A/Accession: T42567
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-3534 <TEL>
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AC59539.1; PID:g2605967
 A/Experimental source: strain NS80567
 C/Genetics:
 A/Note: 24
 C/Superfamily: varicella-zoster virus gene 22 protein

Query Match 16.6%; Score 76; DB 2; Length 3534;
 Best Local Similarity 30.2%; Pred. No. 69;
 Matches 29; Conservative 6; Mismatches 39; Indels 22; Gaps 5;

QY 2 NMSQPVSNVRAIQANINIPMGAFRPGAGPP--RRKCT--PEVEGVP-PTSDEKK 55
 DB 2576 NTKQPPDNRLPVSENNLP--HFVQTPAPPTDTSKPCTVIGSQONLGTAPQKEPKK 2633
 QY 56 PIPGAKKLPG-----PAVNLSEIQNI 77
 DB 2634 PTNNASTAVGSTNKTDTDEPQVQPPSKVASEANNIK 2669

RESULT 13
 A37127
 microtubule-associated protein U - bovine
 N/Alternate names: 190K microtubule-associated protein, MAP-U
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change 05-Nov-1999
 C/Accession: A37127; A32866
 R/Alzawa, H.; Emori, Y.; Murofushi, H.; Kawasaki, H.; Sakai, H.; Suzuki, K.
 J. Biol. Chem. 265, 13849-13855, 1990
 A/Title: Molecular cloning of a ubiquitously distributed microtubule-associated protein
 A/Reference number: A37127; MUID:90338002; PMID:2380192
 A/Accession: A37127
 A/Molecule type: mRNA
 A/Residues: 1-1072 <A1Z>
 A/Cross-references: GB:D90149; GB:J05557; NID:g217577; PIDN:BA14179.1; PID:dt014882; PT
 R:Alzawa, H.; Kawasaki, H.; Murofushi, H.; Kotani, S.; Suzuki, K.; Sakai, H.
 J. Biol. Chem. 264, 5885-5890, 1989

A/Title: A common amino acid sequence in 190-kDa microtubule-associated protein and tau
 A/Reference number: A32866; MUID:89174649; PMID:2494169
 A/Accession: A32866
 A/Molecule type: protein
 A/Residues: 880-901 <A1Z>
 C/Superfamily: MAP2/tau repeat homology
 C/Keywords: microtubule binding; tandem repeat
 F/877-907/Domain: MAP2/tau repeat homology <MT1>
 F/880-901/Region: microtubule binding #status predicted
 F/933-945/Domain: MAP2/tau repeat homology #status atypical <MT2>
 F/946-976/Domain: MAP2/tau repeat homology <MT3>
 F/977-1004/Domain: MAP2/tau repeat homology #status atypical <MT4>

Query Match 16.4%; Score 75.5; DB 2; Length 1072;
 Best Local Similarity 27.4%; Pred. No. 21;
 Matches 23; Conservative 18; Mismatches 36; Indels 7; Gaps 4;

QY 2 NMSQPVSNVRAIQANINIPMGAFRPGAGPPRRKCTPEVEGVPPTSD--EKKIP 58
 DB 804 DLSRKSSTTSVKKSTTVGTA--PPAGAPRARPRATPRSGTPVDKKTPTAKPTS 861
 QY 59 GAKKLPGAVNL--EIQNIKSEL 80
 DB 862 SAPRLGRVANAASAPDLKNVRSKY 885

RESULT 14
 T113065
 PiP2 protein - fruit fly (Drosophila melanogaster)
 C/Species: Drosophila melanogaster
 C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C/Accession: T113065
 R/Suri, V.; Qian, Z.; Hall, J.C.; Roebash, M.
 Neuron 21, 225-234, 1998
 A/Title: Evidence that TIM light response is relevant to light-induced phase shifts in
 A/Reference number: Z17597; MUID:9831331; PMID:9697866
 A/Accession: T113065
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1171 <SUR>
 A/Cross-references: EMBL:AF067153; NID:g3171244; PID:g3171245; PIDN:AA018395.1
 A/Experimental source: strain Canton-S, photoreceptor
 C/Genetics:
 A/Cross-references: FlyBase:FBgn0024943

Query Match 16.2%; Score 74.5; DB 2; Length 1171;
 Best Local Similarity 34.1%; Pred. No. 29;
 Matches 15; Conservative 6; Mismatches 16; Indels 7; Gaps 1;

QY 31 QPPRRKCTPEVEGVPPTSDDEKKP-----IPGAKKLPGRA 67
 DB 886 RPPYKRPATPPIDAVPSPVSKPPPPSVETIPVSLPSPA 929

RESULT 15
 S37671
 MHC class III histocompatibility antigen HLA-B-associated protein 2 [similarity] - huma
 C/Species: Homo sapiens (man)
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 15-Sep-2000
 C/Accession: S37671
 R/Bougueleret, L.
 Submitted to the EMBL Data Library, August 1992
 A/Reference number: S37671
 A/Accession: S37671
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1870 <BOU>
 A/Cross-references: EMBL:Z15025; NID:g29374; PID:g29375
 C/Genetics:
 A/Map position: 6p21.3
 A/Intons: 38/2; 97/2; 129/3; 154/1; 202/1; 252/3; 279/2; 327/1; 357/2; 429/3; 588/1; 6
 C/Superfamily: collagen alpha 1(IV) chain

Query Match 16.1%; Score 74; DB 2; Length 1870;
 Best Local Similarity 35.0%; Pred. No. 54;
 Matches 21; Conservative 8; Mismatches 15; Indels 16; Gaps 4;
 QY 19 NIPMGAFRPGAGOPPRRKECTPVEBEG-VPPTDDEKK-----PIPAKK-----LPGP 66
 Db 1158 SLPEGALSPG-----PRRREAPFOVCFGWSPPAKSLAPKKPPTGTPLPSPSKKPLKELIPGP 1213

Search completed: April 9, 2003, 12:39:40
 Job time : 26.3548 secs

GenCore version 5.1.4 ps_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:33:56 ; Search time 14.2137 Seconds

(without alignments)
155.253 Million cell updates/sec

Title: US-09-647-019-5

Perfect score: 390

Sequence: 1 MSKOPASNSINSIQANINIPM.....KKPIPAVKLPGPAFLSET 75

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.rep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.rep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.rep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.rep.*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.rep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.rep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.5	19.1	244	1 US-08-696-827-1	Sequence 1, App11
2	67.5	17.3	1235	2 US-08-680-326-36	Sequence 36, App1
3	66.5	17.1	520	6 5223391-5	Patent No. 5223391
4	66	16.9	232	4 US-09-149-476-623	Sequence 623, App
5	65	16.7	449	2 US-08-839-008-2	Sequence 2, App11
6	65	16.7	449	2 US-08-839-008-9	Sequence 2, App11
7	65	16.7	1274	4 US-09-095-443-2	Sequence 9, App11
8	63.5	16.3	180	6 5273901-7	Patent No. 5273901
9	63.5	16.3	180	6 5482709-6	Patent No. 5482709
10	63.5	16.3	941	4 US-07-757-0228-14	Sequence 14, App1
11	63.5	16.3	1022	4 US-07-757-0228-84	Sequence 84, App1
12	63.5	16.3	1038	4 US-07-757-0228-74	Sequence 74, App1
13	63.5	16.3	1049	4 US-07-757-0228-58	Sequence 58, App1
14	63.5	16.3	1070	3 US-08-992-635-22	Sequence 22, App1
15	63.5	16.3	1140	4 US-07-757-0228-104	Sequence 104, App1
16	63.5	16.3	1270	4 US-07-757-0228-44	Sequence 44, App1
17	63.5	16.3	1311	4 US-07-757-0228-42	Sequence 42, App1
18	63.5	16.3	1313	4 US-07-757-0228-142	Sequence 142, App
19	63.5	16.3	1314	4 US-07-757-0228-50	Sequence 50, App1
20	63.5	16.3	1330	4 US-07-757-0228-46	Sequence 46, App1
21	63.5	16.3	1330	4 US-07-757-0228-60	Sequence 60, App1
22	63.5	16.3	1354	4 US-07-757-0228-40	Sequence 40, App1
23	63.5	16.3	1361	4 US-07-757-0228-52	Sequence 52, App1
24	63.5	16.3	1404	4 US-07-757-0228-2	Sequence 2, App11
25	63.5	16.3	1404	4 US-07-757-0228-62	Sequence 62, App11
26	63.5	16.3	1404	4 US-07-757-0228-62	Sequence 62, App11
27	63	16.2	489	2 US-08-794-795-7	Sequence 7, App11

28	63	16.2	489	4 US-09-249-200-7	Sequence 7, App11
29	63	16.2	518	1 US-08-392-367B-2	Sequence 2, App11
30	63	16.2	518	3 US-08-893-67B-2	Sequence 2, App11
31	63	16.2	905	2 US-08-574-959A-9	Sequence 9, App11
32	63	16.2	905	4 US-09-357-014-9	Sequence 9, App11
33	63	16.2	1135	2 US-08-574-959A-7	Sequence 7, App11
34	63	16.2	1135	4 US-09-357-014-7	Sequence 7, App11
35	62.5	16.0	83	2 US-08-499-676A-10	Sequence 10, App1
36	62	15.9	159	4 US-08-483-533-28	Sequence 28, App1
37	62	15.9	159	4 US-09-283-871A-28	Sequence 28, App1
38	62	15.9	355	4 US-08-483-533-41	Sequence 41, App1
39	62	15.9	355	4 US-09-283-871A-41	Sequence 41, App1
40	62	15.9	355	5 PCT-US91-06532-3	Sequence 3, App11
41	62	15.9	1400	1 US-08-080-255-7	Sequence 7, App11
42	62	15.9	1400	3 US-08-465-713-7	Sequence 7, App11
43	62	15.9	1400	5 PCT-US93-05857-7	Sequence 7, App11
44	62	15.9	3969	4 US-08-061-376-5	Sequence 5, App11
45	61.5	15.8	179	4 US-09-612-126-11	Sequence 11, App1

ALIGNMENTS

RESULT 1
US-08-696-827-1
Sequence 1, Application US/08696827
Patent No. 5798213
GENERAL INFORMATION:
APPLICANT: MIYADERA Kazutaka
APPLICANT: YAMADA, Yuji
APPLICANT: TAKEBAYASHI, Yuji
APPLICANT: AKIYAMA, Shinichi
TITLE OF INVENTION: MONOCLONAL ANTIBODIES
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSER: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W., Suite 800
CITY: Washington, D.C.
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,827
FILING DATE: 21-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/USJP/02661
FILING DATE: 25-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 327328/1994
FILING DATE: 28-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: Q-42514
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7660
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 244 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-696-827-1
Query Match 19.1%, Score 74.5, DB 1, Length 244;

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22

EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,589
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,593
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,614
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,578
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,576
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/047,501
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,670
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/056,632
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,664
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,876
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,881
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,909
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,875
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,862
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,887
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,908
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/057,650
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/056,884
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,669
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/049,610
EARLIER FILING DATE: 1997-06-13
EARLIER APPLICATION NUMBER: 60/061,060
EARLIER FILING DATE: 1997-10-02

Query Match 16.9%; Score 66; DB 4; Length 232;
Best Local Similarity 28.6%; Pred. No. 3.6;
Matches 10; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

OY 26 GAGOPKPKRSTEEHQHVPTPESEKSEKKPIF 60
DB 166 GSGLPWESQFOGEDRKFAPBDKSPPTTEREQVP 200

RESULT 5

US-08-839-008-2

Sequence 2, Application US/08839008

Patent No. 5916758

GENERAL INFORMATION:

APPLICANT: Hurle, Mark R

APPLICANT: McDonnell, Peter C

APPLICANT: McNulty, Dean E

APPLICANT: Rosen, Craig A

APPLICANT: Siemens, Ivo R

APPLICANT: Young, Peter R

APPLICANT: Yue, Tian-Li

TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/839,008

FILING DATE: 23-APR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/563,697

FILING DATE: 28-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Baumeister, Kirk

REGISTRATION NUMBER: 33,833

REFERENCE/DOCKET NUMBER: P50384

TELEPHONE: 610-270-5096

TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 449 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-839-008-2

Query Match 16.7%; Score 65; DB 2; Length 449;

Best Local Similarity 37.3%; Pred. No. 11;

Matches 19; Conservative 7; Mismatches 19; Indels 6; Gaps 3;

OY 18 IPMGAFRPGAGOPPKRKESTEEHQHVPTPESE--EKSEKKPIGAVKLP 66

DB 273 LPRGTAKGQGPGRK--GTEPKVLP-PKSGPPEKTESPAPAPCP 319

RESULT 6

US-08-839-008-9

Sequence 9, Application US/08839008

Patent No. 5916758

GENERAL INFORMATION:

APPLICANT: Hurle, Mark R

APPLICANT: McDonnell, Peter C

APPLICANT: McNulty, Dean E

APPLICANT: Rosen, Craig A

APPLICANT: Siemens, Ivo R

APPLICANT: Young, Peter R

APPLICANT: Yue, Tian-Li

TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/839,008

FILING DATE: 23-APR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/563,697

FILING DATE: 28-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Baumeister, Kirk

REGISTRATION NUMBER: 33,833

REFERENCE/DOCKET NUMBER: P50384

TELEPHONE: 610-270-5096

TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 449 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-839-008-9

Query Match 16.7%; Score 65; DB 2; Length 449;

Best Local Similarity 37.3%; Pred. No. 11;

Matches 19; Conservative 7; Mismatches 19; Indels 6; Gaps 3;

OY 18 IPMGAFRPGAGOPPKRKESTEEHQHVPTPESE--EKSEKKPIGAVKLP 66

DB 273 LPRGTAKGQGPGRK--GTEPKVLP-PKSGPPEKTESPAPAPCP 319

RESULT 7

US-09-095-443-2

Sequence 2, Application US/09095443

Patent No. 6342593

GENERAL INFORMATION:

APPLICANT: Plowman, Gregory

APPLICANT: Peles, Eior

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lyon & Lyon

STREET: 633 West Fifth Street

CITY: Suite 4700

STATE: Los Angeles

COUNTRY: California

ZIP: U.S.A.

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: storage

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: FastSeq for Windows 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/095,443
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/049,477
FILING DATE: June 12, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 235/055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1274 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-095-443-2

Query Match 16.3%; Score 65; DB 4; Length 1274;
Best Local Similarity 28.2%; Pred. No. 39;
Matches 24; Conservative 13; Mismatches 30; Indels 18; Gaps 4;
OY 5 PASNRSIOANINI-PMG-----AFRPGAGQPPKKESTEEBQHVPT-----P 47
DB 1132 FISSIOATIAKLSIRPGLESFVASLPQAPFPGLPRASLESTIPFSSSPPLSLP 1191
OY 48 EEESESEKKPIPGAVKLPGPAPNL 72
DB 1192 EAPQKER-PPVPEAPSSGPPSSSL 1215

RESULT 8
5273901-7
PATENT NO. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581,693
FILING DATE: 12-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 7
LENGTH: 180
5273901-7

Query Match 16.3%; Score 63.5; DB 6; Length 180;
Best Local Similarity 42.1%; Pred. No. 5.1;
Matches 16; Conservative 3; Mismatches 16; Indels 3; Gaps 1;
OY 32 KRKEFSTEEBQHVPTPESESEKKPIPGAVKLPGPA 69
DB 11 KRRE---EEERBERGEEKEERBAAPAAATTAAPA 45

RESULT 9
5482709-6
PATENT NO. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;

AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 6
LENGTH: 180
5482709-6

Query Match 16.3%; Score 63.5; DB 6; Length 180;
Best Local Similarity 42.1%; Pred. No. 5.1;
Matches 16; Conservative 3; Mismatches 16; Indels 3; Gaps 1;
OY 32 KRKEFSTEEBQHVPTPESESEKKPIPGAVKLPGPA 69
DB 11 KRRE---EEERBERGEEKEERBAAPAAATTAAPA 45

RESULT 10
US-07-757-022B-14
Sequence 14, Application US/07/757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 941 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-14

Query Match 16.3%; Score 63.5; DB 4; Length 941;
Best Local Similarity 28.4%; Pred. No. 40;
Matches 19; Conservative 7; Mismatches 40; Indels 1; Gaps 1;

QY 3 KQPA-SNIRSIQANINIPMGAFRPGAGPPKRFSTEEBOHVPTPESEKSEKKPIPG 61
DB 321 KEPAPTTKSAPTTKKEPAPTTKSAPTTKPEPSPTTKKEPAPTTKKEPAPTTKKEPAPT 380

QY 62 AVKLPGP 68
DB 381 TPKEPAP 387

RESULT 11
US-07-757-022B-84
Sequence 84, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: AMINO ACID

TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-84

Query Match 16.3%; Score 63.5; DB 4; Length 1022;
Best Local Similarity 28.4%; Pred. No. 44;
Matches 19; Conservative 7; Mismatches 40; Indels 1; Gaps 1;

QY 3 KQPA-SNIRSIQANINIPMGAFRPGAGPPKRFSTEEBOHVPTPESEKSEKKPIPG 61
DB 402 KEPAPTTKSAPTTKKEPAPTTKSAPTTKPEPSPTTKKEPAPTTKKEPAPTTKKEPAPT 461

QY 62 AVKLPGP 68
DB 462 TPKEPAP 468

RESULT 12
US-07-757-022B-74
Sequence 74, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Gesner, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-74
Query Match 16.3%; Score 63.5; DB 4; Length 1038;
Best Local Similarity 28.4%; Pred. No. 45;

Matches 19; Conservative 7; Mismatches 40; Indels 1; Gaps 1;
QY 3 KQPA-SNIRSIQANINIPMGAFRPGAGOPPKRKEFSTEEHQHVPTESEKSEKPIPG 61
DB 386 KEPAFTTKSAPTTKKEPAFTTKSAPTTKKEPSPTTKKEPAFTTKKEPAFTTKKEPAFT 445
QY 62 AVKLPGP 68
DB 446 TPKEPAP 452

RESULT 13

US-07-757-022B-58
Sequence 58, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Geert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-58

Query Match 16.3%; Score 63.5; DB 4; Length 1049;
Best Local Similarity 28.4%; Pred. No. 45;

Matches 19; Conservative 7; Mismatches 40; Indels 1; Gaps 1;

QY 3 KQPA-SNIRSIQANINIPMGAFRPGAGOPPKRKEFSTEEHQHVPTESEKSEKPIPG 61
DB 429 KEPAFTTKSAPTTKKEPAFTTKSAPTTKKEPSPTTKKEPAFTTKKEPAFTTKKEPAFT 488

QY 62 AVKLPGP 68
DB 489 TPKEPAP 495

RESULT 14

US-08-922-635-22
Sequence 22, Application US/08922635A
Patent No. 603871
GENERAL INFORMATION:
APPLICANT: PILETZ, John E.
APPLICANT: IVANOV, Tina R.
TITLE OF INVENTION: DNA MOLECULES ENCODING IMIDALINE RECEPTIVE POLYPEPTIDES
FILE REFERENCE: Corrected Sequence Listing
Patent No. 603871
CURRENT APPLICATION NUMBER: US/08/922,635A
CURRENT FILING DATE: 1997-09-03
EARLIER APPLICATION NUMBER: 08/650,766
EARLIER FILING DATE: 1996-05-20
EARLIER APPLICATION NUMBER: 60/012,600
EARLIER FILING DATE: 1996-03-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 22
LENGTH: 1070
TYPE: PRT
ORGANISM: Homo sapiens
US-08-922-635-22

Query Match 16.3%; Score 63.5; DB 3; Length 1070;
Best Local Similarity 33.3%; Pred. No. 47;
Matches 17; Conservative 6; Mismatches 19; Indels 9; Gaps 2;

QY 19 PMGAFRPGAGOPPKRKEFSTEEHQHVPTESEKSEKPIPGAVTLPGPA 69
DB 589 POGSFF-ADGQPAERASNDQRPQEVF-----AALAPAVGVAPAPAP 630

RESULT 15

US-07-757-022B-104
Sequence 104, Application US/07757022B
Patent No. 6433142
GENERAL INFORMATION:
APPLICANT: Geener, Thomas G.
APPLICANT: Clark, Stephen C.
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseer, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 1140 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-104

Query Match 16.3%; Score 63.5; DB 4; Length 1140;
Best Local Similarity 28.4%; Pred. No. 50;
Matches 19; Conservative 7; Mismatches 40; Indels 1; Gaps 1;
QY 3 KQPA-SNIRSIQANINIPMGAFRPGAGOPPKRKEFSTEEQHVPTPESEKSEKKPIIG 61
DB 520 KEPAPTTKSAPTTKBPAPTTKSAPTTPKKEPSPPTTKBPAPTTKBPAPTTKBPAPT 579
QY 62 AVKLPGP 68
DB 580 TPKEPAP 586

Search completed: April 9, 2003, 12:40:32
Job time : 15.2137 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2003, 12:38:46 ; Search time 16.6331 Seconds

(without alignments)
275.668 Million cell updates/sec

Title: US-09-647-019-5

Perfect score: 390

Sequence: 1 MSKQPSNIRSIQANINIPM.....KKPIGAVKLPGPANLSET 75

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

Total number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodaca/2/pubpaa/US08_NEW_PUB_PEP.*
- 2: /cgn2_6/ptodaca/2/pubpaa/PTCT_NEW_PUB_PEP.*
- 3: /cgn2_6/ptodaca/2/pubpaa/US06_NEW_PUB_PEP.*
- 4: /cgn2_6/ptodaca/2/pubpaa/US06_PUBCOMB_PEP.*
- 5: /cgn2_6/ptodaca/2/pubpaa/US07_NEW_PUB_PEP.*
- 6: /cgn2_6/ptodaca/2/pubpaa/US07_PUBCOMB_PEP.*
- 7: /cgn2_6/ptodaca/2/pubpaa/PTCTUS_PUBCOMB_PEP.*
- 8: /cgn2_6/ptodaca/2/pubpaa/US08_PUBCOMB_PEP.*
- 9: /cgn2_6/ptodaca/2/pubpaa/US09_NEW_PUB_PEP.*
- 10: /cgn2_6/ptodaca/2/pubpaa/US09_PUBCOMB_PEP.*
- 11: /cgn2_6/ptodaca/2/pubpaa/US10_NEW_PUB_PEP.*
- 12: /cgn2_6/ptodaca/2/pubpaa/US10_PUBCOMB_PEP.*
- 13: /cgn2_6/ptodaca/2/pubpaa/US60_NEW_PUB_PEP.*
- 14: /cgn2_6/ptodaca/2/pubpaa/US60_PUBCOMB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	18.7	603	US-09-764-868-705	Sequence 705, App
2	72.5	18.6	1235	US-09-904-065-8	Sequence 8, Appl
3	72.5	18.6	1235	US-09-904-065-19	Sequence 19, Appl
4	69	17.7	930	US-10-113-794A-1	Sequence 1, Appl
5	68	17.4	260	US-10-102-806-503	Sequence 503, App
6	67	17.2	300	US-10-278-173-34	Sequence 34, Appl
7	66.5	17.1	1235	US-09-904-065-6	Sequence 6, Appl
8	66.5	17.1	1235	US-09-904-065-16	Sequence 16, Appl
9	66.5	17.1	1235	US-09-904-065-17	Sequence 17, Appl
10	66	16.9	232	US-09-809-391-623	Sequence 623, App
11	65.5	16.8	1690	US-09-788-043C-5	Sequence 5, Appl
12	65	16.7	449	US-09-919-497-89	Sequence 89, Appl
13	65	16.7	458	US-09-925-301-1282	Sequence 1282, Ap
14	65	16.7	603	US-09-906-779-4	Sequence 4, Appl
15	65	16.7	617	US-09-864-761-36182	Sequence 36182, A
16	65	16.7	1274	US-10-020-215-2	Sequence 2, Appl
17	64.5	16.7	2665	US-09-864-761-34248	Sequence 34248, A
18	64.5	16.5	360	US-10-149-819-6	Sequence 10, Appl
19	64.5	16.5	1235	US-09-904-065-10	Sequence 10, Appl

20	64.5	16.5	1235	US-09-904-065-18	Sequence 18, Appl
21	64	16.4	256	US-09-789-054A-40	Sequence 40, Appl
22	63.5	16.3	941	US-10-124-557-14	Sequence 14, Appl
23	63.5	16.3	1022	US-10-124-557-84	Sequence 84, Appl
24	63.5	16.3	1038	US-10-124-557-74	Sequence 74, Appl
25	63.5	16.3	1049	US-10-124-557-58	Sequence 58, Appl
26	63.5	16.3	1140	US-10-124-557-104	Sequence 104, App
27	63.5	16.3	1270	US-10-124-557-44	Sequence 44, Appl
28	63.5	16.3	1311	US-10-124-557-42	Sequence 42, Appl
29	63.5	16.3	1313	US-10-124-557-142	Sequence 142, Appl
30	63.5	16.3	1314	US-10-124-557-50	Sequence 50, Appl
31	63.5	16.3	1320	US-10-124-557-46	Sequence 46, Appl
32	63.5	16.3	1320	US-10-124-557-60	Sequence 60, Appl
33	63.5	16.3	1354	US-10-124-557-48	Sequence 48, Appl
34	63.5	16.3	1361	US-10-124-557-40	Sequence 40, Appl
35	63.5	16.3	1363	US-10-124-557-52	Sequence 52, Appl
36	63.5	16.3	1404	US-10-124-557-2	Sequence 2, Appl
37	63.5	16.3	1404	US-10-124-557-62	Sequence 62, Appl
38	63.5	16.3	1504	US-09-932-145-7	Sequence 7, Appl
39	62.5	16.0	365	US-10-078-929-164	Sequence 164, App
40	62.5	16.0	428	US-09-922-331-16	Sequence 16, Appl
41	62	15.9	112	US-09-867-550-1018	Sequence 1018, Ap
42	62	15.9	255	US-09-992-598-287	Sequence 287, App
43	62	15.9	255	US-09-989-293A-287	Sequence 287, App
44	62	15.9	255	US-09-989-735-287	Sequence 287, App
45	62	15.9	255	US-09-990-444-287	Sequence 287, App

ALIGNMENTS

RESULT 1

US-09-764-868-705

Sequence 705, Application US/09764868

Patent No. US20020168711A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT232

CURRENT APPLICATION NUMBER: US/09/764,868

CURRENT FILING DATE: 2001-01-17

Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 705

LENGTH: 603

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (163)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-868-705

Query Match

Best Local Similarity 18.7%; Score 73; DB 9; Length 603;

Matches 25; Conservative 12; Mismatches 35; Indels 30; Gaps 3;

QY 2 SKQPSNIRSIQANINIPGAFRPGAGOPPKKEKST-----EEQHVP----- 45

DB 112 SRPFRSDSKLTSRDSGPGFGSGSYHGRKRSKTCRQGRSVSSQBEKSSPVKAS 171

QY 46 -----TPE-----SREKSEKKPIGAVKLPGPANLS 73

DB 172 VKRTMPDKTATPERPAPENAPSSKILPAPDKVSPKTLT 213

RESULT 2

US-09-904-065-8

Sequence 8, Application US/09904065

Patent No. US20020076789A1

GENERAL INFORMATION:

APPLICANT: Homa, Fred

```

; APPLICANT: Mathen, Michael
; APPLICANT: Hopkins, Todd
; APPLICANT: Thomsen, Darrell
; TITLE OF INVENTION: A Method for Treating Herpes Virus
; FILE REFERENCE: 00221
; CURRENT APPLICATION NUMBER: US/09/904,065
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: herpes simplex
; US-09-904-065-8

```

```

Query Match          18.6%; Score 72.5; DB 10; Length 1235;
Best Local Similarity 34.8%; Pred. No. 22;
Matches 16; Conservative 9; Mismatches 18; Indels 3; Gaps 1;

```

```

QY      21 GAFPRGAGOPPKRKKEFTSEBQHVPPTPESEKSEKKPIPGAVKLP 66
DB      641 GRFRGGGGGAPRRPAAREDEER---PSEEGEDDEREGGGEREP 683

```

```

RESULT 3
; US-09-904-065-19
; Sequence 19, Application US/09904065
; Patent No. US20020076789A1
; GENERAL INFORMATION:
; APPLICANT: Homa, Fred
; APPLICANT: Mathen, Michael
; APPLICANT: Hopkins, Todd
; APPLICANT: Thomsen, Darrell
; TITLE OF INVENTION: A Method for Treating Herpes Virus
; FILE REFERENCE: 00221
; CURRENT APPLICATION NUMBER: US/09/904,065
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 19
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: herpes simplex
; US-09-904-065-19

```

```

Query Match          18.6%; Score 72.5; DB 10; Length 1235;
Best Local Similarity 34.8%; Pred. No. 22;
Matches 16; Conservative 9; Mismatches 18; Indels 3; Gaps 1;

```

```

QY      21 GAFPRGAGOPPKRKKEFTSEBQHVPPTPESEKSEKKPIPGAVKLP 66
DB      641 GRFRGGGGGAPRRPAAREDEER---PSEEGEDDEREGGGEREP 683

```

```

RESULT 4
; US-10-113-794A-1
; Sequence 1, Application US/10113794A
; Publication No. US20030022202A1
; GENERAL INFORMATION:
; APPLICANT: Flanagan et al.
; TITLE OF INVENTION: B EPHRIN REGULATION OF G-PROTEIN COUPLED
; TITLE OF INVENTION: CHEMOTACTRACION
; FILE REFERENCE: 2535/106
; CURRENT APPLICATION NUMBER: US/10/113,794A
; CURRENT FILING DATE: 2002-04-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-113-794A-1

```

```

Query Match          17.7%; Score 69; DB 9; Length 930;
Best Local Similarity 38.3%; Pred. No. 36;
Matches 18; Conservative 6; Mismatches 17; Indels 6; Gaps 2;

```

```

QY      27 AGOPPKRKKEFSTE-BEQHVP-----PSESEKSEKKPIPGAVKLP 67
DB      391 AAPPTKRMETEADEKEMPLVGEKGPAGAEPAAPSXNPSFGQLPFG 437

```

```

RESULT 5
; US-10-102-806-503
; Sequence 503, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 503
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (69)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-102-806-503

```

```

Query Match          17.4%; Score 68; DB 9; Length 260;
Best Local Similarity 28.6%; Pred. No. 11;
Matches 20; Conservative 8; Mismatches 24; Indels 18; Gaps 2;

```

```

QY      4 QPASNIRSIQANINIPGAFPRGAGOPPKRKKEFTSEB-QHVP----- 47
DB      66 QPQVFPDSVNARILPLPAEYFSGVQLPPLSPFTYERKGDVPEPKILALRGEDPGN 125

```

```

QY      48 --ESEKSEE 55.
DB      126 LNESEEEEE 135

```

```

RESULT 6
; US-10-278-173-34
; Sequence 34, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiaang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddie, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: M1-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392

```

```

; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1004
US-10-278-173-34

```

```

Query Match      17.1%; Score 67; DB 9; Length 300;
Best Local Similarity 24.7%; Pred. No. 16;
Matches 18; Conservative 19; Mismatches 32; Indels 4; Gaps 2;

```

```

QY 4 QPANSIRIQNINIPM--GAPRPGAGPPPKKSTEEBOHVPTPESEKSEKKPIPG 61
DB 198 EAARLRGSKAILNPFLEVGMKPRADGEEKRK--RDDDKVTVEKVLKTEOSVDVNG 255

```

```

QY 62 AVKLPGAPNLSE 74
DB 256 GETPPTVTSNITE 268

```

```

RESULT 7
US-09-904-065-6
; Sequence 6, Application US/09904065
; Patent No. US20020076789A1
; GENERAL INFORMATION:
; APPLICANT: Homa, Fred
; APPLICANT: Mathen, Michael
; APPLICANT: Hopkins, Todd
; APPLICANT: Thomsen, Darrell
; TITLE OF INVENTION: A Method for Treating Herpes Virus
; FILE REFERENCE: 00221
; CURRENT APPLICATION NUMBER: US/09/904,065
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: herpes simplex
US-09-904-065-6

```

```

Query Match      17.1%; Score 66.5; DB 10; Length 1235;
Best Local Similarity 32.6%; Pred. No. 89;
Matches 15; Conservative 9; Mismatches 19; Indels 3; Gaps 1;

```

```

QY 21 GAFRPGAGPPPKKSTEEBOHVPTPESEKSEKKPIPGAVKL 66
DB 641 GFRFGAGGAPKRPAAAREDEER--PEEGEDEDEREGGGEREP 683

```

```

RESULT 8
US-09-904-065-16
; Sequence 16, Application US/09904065
; Patent No. US20020076789A1
; GENERAL INFORMATION:
; APPLICANT: Homa, Fred
; APPLICANT: Mathen, Michael
; APPLICANT: Hopkins, Todd
; APPLICANT: Thomsen, Darrell
; TITLE OF INVENTION: A Method for Treating Herpes Virus
; FILE REFERENCE: 00221
; CURRENT APPLICATION NUMBER: US/09/904,065
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 1235

```

```

; TYPE: PRT
; ORGANISM: herpes simplex
US-09-904-065-16

```

```

Query Match      17.1%; Score 66.5; DB 10; Length 1235;
Best Local Similarity 32.6%; Pred. No. 89;
Matches 15; Conservative 9; Mismatches 19; Indels 3; Gaps 1;

```

```

QY 21 GAFRPGAGPPPKKSTEEBOHVPTPESEKSEKKPIPGAVKL 66
DB 641 GFRFGAGGAPKRPAAAREDEER--PEEGEDEDEREGGGEREP 683

```

```

RESULT 9
US-09-904-065-17
; Sequence 17, Application US/09904065
; Patent No. US20020076789A1
; GENERAL INFORMATION:
; APPLICANT: Homa, Fred
; APPLICANT: Mathen, Michael
; APPLICANT: Hopkins, Todd
; APPLICANT: Thomsen, Darrell
; TITLE OF INVENTION: A Method for Treating Herpes Virus
; FILE REFERENCE: 00221
; CURRENT APPLICATION NUMBER: US/09/904,065
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 1235
; TYPE: PRT
; ORGANISM: herpes simplex
US-09-904-065-17

```

```

Query Match      17.1%; Score 66.5; DB 10; Length 1235;
Best Local Similarity 32.6%; Pred. No. 89;
Matches 15; Conservative 9; Mismatches 19; Indels 3; Gaps 1;

```

```

QY 21 GAFRPGAGPPPKKSTEEBOHVPTPESEKSEKKPIPGAVKL 66
DB 641 GFRFGAGGAPKRPAAAREDEER--PEEGEDEDEREGGGEREP 683

```

```

RESULT 10
US-09-809-391-623
; Sequence 623, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 623
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (8)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (9)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (11)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

```

NAME/KEY: SITE
LOCATION: (12)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (232)
OTHER INFORMATION: Xaa equals stop translation
US-09-809-391-623

Query Match 16.9%; Score 66; DB 9; Length 232;
Best Local Similarity 28.6%; Pred. No. 15;
Matches 10; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

QY 26 GAGPPKPKRSTEEHGVPTPESEKSEKPP 60
DB 166 GSGLPPEMSPQGGEDRKAPSDKSPPTTREQVP 200

RESULT 11
US-09-788-043C-5
Sequence 5, Application US/09788043C
Patent No. US20020107361A1
GENERAL INFORMATION:
APPLICANT: Heller, Renu
APPLICANT: Zuo, Fengrong
APPLICANT: Klonowski, Paul
TITLE OF INVENTION: No. US20020107361A1 Metalloproteases Having
TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
FILE REFERENCE: ROCH-004
CURRENT APPLICATION NUMBER: US/09/788, 043C
CURRENT FILING DATE: 2001-02-16
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: 60/184,152
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 1690
TYPE: PRT
ORGANISM: human
US-09-788-043C-5

Query Match 16.8%; Score 65.5; DB 10; Length 1690;
Best Local Similarity 28.9%; Pred. No. 1.6e+02;
Matches 24; Conservative 10; Mismatches 32; Indels 17; Gaps 5;

QY 2 SKOPASNIRSIQANINIPMGAFR-GAGPPKPKRSTEE---QHV---PTPESEKRS- 53
DB 984 AQQPASEVTCSPLCRWPLGTGPEGSSGSSSHLFNEADPIPHHLPAPSPASPKRG 1043

QY 54 -----EKKPPLPGAVKLPKPAF 70
DB 1044 TMGNAIEEAR---ELDLPGVF 1063

RESULT 12
US-09-919-497-89
Sequence 89, Application US/09919497
Patent No. US2002010662A1
GENERAL INFORMATION:
APPLICANT: Muller, George L.
TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
FILE REFERENCE: B0801/7225
CURRENT APPLICATION NUMBER: US/09/919,497
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/221,735
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn version 3.0
SEQ ID NO 89
LENGTH: 449
TYPE: PRT
ORGANISM: Homo sapiens
US-09-919-497-89

Query Match 16.7%; Score 65; DB 10; Length 449;
Best Local Similarity 37.3%; Pred. No. 41;
Matches 19; Conservative 7; Mismatches 19; Indels 6; Gaps 3;

QY 18 IPMGAPRPGAGPPKPKRSTEEHGVPTPESE--EKSEKRP1PGAVKLP 66
DB 273 LPRGTAKGGQGPGRK---GTEPRVKLP-PKSGPPKRTESPSAPDAPTCP 319

RESULT 13
US-09-925-301-1282
Sequence 1282, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1282
LENGTH: 458
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (249)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1282

Query Match 16.7%; Score 65; DB 10; Length 458;
Best Local Similarity 37.3%; Pred. No. 41;
Matches 19; Conservative 7; Mismatches 19; Indels 6; Gaps 3;

QY 18 IPMGAPRPGAGPPKPKRSTEEHGVPTPESE--EKSEKRP1PGAVKLP 66
DB 282 LPRGTAKGGQGPGRK---GTEPRVKLP-PKSGPPKRTESPSAPDAPTCP 328

RESULT 14
US-09-906-779-4
Sequence 4, Application US/09906779
Patent No. US2002006484A1
GENERAL INFORMATION:
APPLICANT: Shi et al.
TITLE OF INVENTION: Human Protein Tyrosine Phosphatase Polynucleotides, Polypeptides
FILE REFERENCE: PTO40P1
CURRENT APPLICATION NUMBER: US/09/906,779
CURRENT FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US01/01563
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/176,306
PRIOR FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 603
TYPE: PRT
ORGANISM: Homo sapiens
US-09-906-779-4

Query Match 16.7%; Score 65; DB 10; Length 603;
Best Local Similarity 28.2%; Pred. No. 57;
Matches 24; Conservative 13; Mismatches 30; Indels 18; Gaps 4;

QY 5 PASNIRSIQANINI-PNG-----APRPGAGPPKPKRSTEEHGVPT-----P 47


```

Db      461 PISIOFTIKLIRPGGLSPVASLPGPAPPGULPSPATPSPSSPPPLSSPLP 520
Qy      48 ESEKSEKKRIPGAVKLPGFAPNL 72
      521 EAPQPKKE-PPVEAPSSGPPSSSL 544

RESULT 15
US-09-864-761-36182
/ Sequence 36182, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aecm1ca-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263,6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 36182
/ LENGTH: 617
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC005529.7
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
/ OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUATE 2.00e-33

```

```

/ OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUATE 3.00e-22
US-09-864-761-36182
Query Match      16.7%; Score 65; DB 10; Length 617;
Best Local Similarity 37.5%; Pred. No 58;
Matches 15; Conservative 7; Mismatches 14; Indels 4; Gaps 1;

Qy      29 QPPKKESTEEQHVPPESEKSEKKRIPGAVKLPGP 68
      434 BPPKK---AEEKCAPATPTEEKOSKKEAPKKAEPK 469

```

Search completed: April 9, 2003, 12:53:42
 Job time : 18.6331 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: April 9, 2003, 12:29:50 ; Search time 26.9153 Seconds
(without alignments)
371.305 Million cell updates/sec

Title: US-09-647-019-5

Perfect score: 390
Sequence: 1 MSKOPASINRISIQNNIPM.....KKPIGAVKLPGAFNLSET 75

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

```
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	390	100.0	75	20	AAV28834
2	274	70.3	88	20	AAV28651
3	274	70.3	88	20	AAE16632
4	262.5	67.3	85	20	AAV28650
5	81	20.8	172	22	ABG11263
6	80.5	20.6	602	22	ABG23750
7	80	20.5	521	22	AAU87304
8	80	20.5	602	23	AAU93023
9	75.5	19.4	222	21	AAAG9497
10	75.5	19.4	1469	22	ABE68168

11	74.5	19.1	167	22	ABG25362	Novel human diagno
12	74.5	19.1	244	17	AA939451	Thymidine phosphor
13	74.5	19.1	323	22	AAU02919	Angiotensin conver
14	74.5	19.1	359	22	AA25759	Human protein sequ
15	74.5	19.1	363	22	AAU02917	Angiotensin conver
16	74.5	19.1	442	22	AAU02918	Angiotensin conver
17	74.5	19.1	482	13	AA20641	Placental-derived
18	74.5	19.1	482	17	AA77531	Angiotensin conver
19	74.5	19.1	482	19	AA62026	Human platelet der
20	74.5	19.1	513	23	ABP41705	Recombinantly prod
21	74.5	19.1	517	22	AAU02976	Human ovarian anfi
22	73.5	18.8	2153	22	AAU33195	Angiotensin conver
23	73	18.7	603	22	AAU87594	Novel human secret
24	73	18.7	603	22	AAU17140	Novel human nerv
25	72.5	18.6	1235	23	AAE18856	Novel signal trans
26	72.5	18.6	1235	23	AAE18855	Herpes simplex vir
27	72	18.5	128	21	AAE18855	Herpes simplex vir
28	71	18.2	356	22	AAE18855	Plasma radlata tran
29	71	18.2	495	22	AAW40566	Human polypeptide
30	71	18.2	893	22	AAW79519	Human polypeptide
31	71	18.2	907	22	AAW78535	Human polypeptide
32	71	18.2	1098	22	ABG20365	Human protein SEQ
33	70	17.9	716	22	ABG20365	Human protein SEQ
34	69	17.7	146	22	ABG29413	Novel human diagno
35	69	17.7	1072	23	ABG57023	Novel human diagno
36	69	17.7	1557	22	ABG63735	Mouse ischaemic co
37	68.5	17.6	672	22	AAW39107	Drosophila melanog
38	68.5	17.6	712	21	AAE43147	Human polypeptide
39	68.5	17.6	739	22	AAW39106	Human polypeptide
40	68.5	17.6	746	22	AAW40892	Human polypeptide
41	68.5	17.6	746	22	AAW40892	Human polypeptide
42	68	17.4	260	21	AAE58795	Human polypeptide
43	68	17.4	280	23	AAU72954	Breast and ovarian
44	68	17.4	280	23	AAU72983	Neisseria meningit
45	68	17.4	771	22	ABE58529	Drosophila melanog

ALIGNMENTS

AAV28834	standard; Protein; 75 AA.
AAV28834	
17-JAN-2000	(first entry)
Xenopus chisel (Cal) protein.	
Chisel protein, Cal; EF-Hand protein super family; muscle development; heart/skeletal muscle cell development; signalling pathway; xenopus; x chromosome; regulation; adaptive process; muscle homeostasis; nucleus; detection; diagnosis; prophylaxis; treatment; skeletal myopathy; muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy; Becker's myotonic dystrophy; heart failure; cardiac hypertrophy; differentiation; gene therapy; transgenic animal; drug screening;	
scallop regulatory myosin light chain.	
Xenopus sp.	
MO9950410-A1.	
07-OCT-1999.	
26-MAR-1999;	99WO-AU00220.
27-MAR-1998;	98AU-0002634.
(CHAN-) CHANG CARDIAC RES INST VICTOR.	
(GEHO) GEN HOSPITAL CORP.	
(HALL-) HALL INST MEDICAL RES WALTER & ELIZA.	

PI Harvey RP, Musaro A, Palmer SJ, Rosenthal NA,
 XX WPI, 1999-610852/52.
 DR
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 11; Fig 3; 157pp; English.
 XX
 CC The present sequence is the xenopus chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 CC
 SQ Sequence 75 AA;
 Query Match 100.0%; Score 390; DB 20; Length 75;
 Best Local Similarity 100.0%; Pred. No. 8.1e-35;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKOPASINRISIQANINIPMGAFRPGAGOPPRKKEFTSEBOHVTPTSESEKSEKKRIP 60
 Db 1 MSKOPASINRISIQANINIPMGAFRPGAGOPPRKKEFTSEBOHVTPTSESEKSEKKRIP 60
 QY 61 GAVKLPGPAFNLSET 75
 Db 61 GAVKLPGPAFNLSET 75
 DE Human chisel (Csl) protein.
 KW Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; regulation;
 KW Xg21.3-q22; adaptive process; muscle homeostasis; cardiac hypertrophy;
 KW detection; diagnosis; prophylaxis; treatment; differentiation; nucleus;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW skeletal myopathy; Becker's myotonic dystrophy; heart failure;
 KW transgenic animal; drug screening; gene therapy; homology;
 KW scallop regulatory myosin light chain.
 OS Homo sapiens.
 PN WO9950410-A1.
 PD 07-OCT-1999.
 PF 26-MAR-1999; 99WO-AU00220.
 PR 27-MAR-1998; 98AU-0002634.
 PA (THAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GHEO) GEN HOSPITAL CORP.

PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA,
 XX WPI, 1999-610852/52.
 DR N-PSDB; AAX90904.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 7; Fig 3; 157pp; English.
 XX
 CC The present sequence is the human chisel protein (Csl), that is a member
 CC of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localised to the nucleus and has 86% homology
 CC to the mouse Csl. Structural homology between Csl and scallop regulatory
 CC myosin light chain is also detected. It is predominantly expressed in
 CC heart and skeletal muscles and is activated after the differentiation of
 CC cells. Csl functions in regulation aspects of differentiation or
 CC adaptive processes that maintain muscle homeostasis. This sequence can
 CC be used in the detection, diagnosis, prophylactic and therapeutic
 CC treatment of diseases such as those involving aberrant muscle cell
 CC development and functional activity. It is also used in the treatment of
 CC muscular and myotonic dystrophies, skeletal myopathies such as Duchenne
 CC muscular dystrophy and Becker's myotonic dystrophy, heart failure,
 CC cardiac hypertrophy, myocarditis, myofiber atrophy, etc. The Csl gene
 CC sequence can also be used in gene therapy, for the production of
 CC transgenic animals and for drug screening.
 CC
 SQ Sequence 88 AA;
 Query Match 70.3%; Score 274; DB 20; Length 88;
 Best Local Similarity 75.7%; Pred. No. 3.3e-22;
 Matches 56; Conservative 6; Mismatches 8; Indels 4; Gaps 2;
 QY 1 MSKOPASINRISIQANINIPMGAFRPGAGOPPRKKEFTSEBOHVTPTSESEKSEKKRIP 60
 Db 3 MSKOPASINRISIQANINIPMGAFRPGAGOPPRKKEFTSEBOHVTPTSESEKSEKKRIP 58
 QY 61 GAVKLPGPAFNLSE 74
 Db 59 GAKKLPGPAVNLSE 72
 DE Human 66214 protein.
 KW Human; congestive heart failure; dilative cardiomyopathy; sudden death;
 KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
 KW heart muscle disease; conduction disorder; coronary heart disease;
 KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
 KW pulmonary heart disease; valvular heart disease; pericardial disease;
 KW congenital heart disease; gene therapy; syncope; transgenic animal;
 KW expressed sequence tag; EST; 66214 protein.
 OS Homo sapiens.
 PN WO200192567-A2.
 PD 06-DEC-2001.
 PF 30-MAY-2001; 2001WO-EP06165.
 PR 30-MAY-2000; 2000US-207400P.
 PA

PA (MEDI-) MEDIGENE AG.
 XX Bunk D, Reuner B, Beck J, Henkel T;
 XX WPI; 2002-122073/16.
 DR N-PSDB; AAD27216.
 XX
 PT Identifying a subject at risk for a heart disease e.g. congestive heart
 PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 PT the polypeptide expressed by genes abnormally expressed in heart tissue
 XX
 PS Claim 1a; Fig 9c; 154pp; English.
 XX
 CC The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present sequence is 66214 protein encoded by an
 CC expressed sequence tag (EST) DNA.
 XX
 SQ Sequence 88 AA;
 Query Match 70.3%; Score 274; DB 23; Length 88;
 Best Local Similarity 75.7%; Pred. No. 3.3e-22;
 Matches 56; Conservative 6; Mismatches 8; Indels 4; Gaps 2;
 QY 1 MSKQPSNIRSIQANINIPMGAFRPGAGOPPRKKEFSTEEBOHVPTPSESEKSEKKRIP 60
 DB 3 MSKQPSNVRAIQANINIPMGAFRPGAGOPPRKKECTEVEEGVP-PTSD--EKKRIP 58
 QY 61 GAVKLRGPAPNLSE 74
 DB 59 GAKKLPGPAPNLSE 72
 RESULT 4
 ID AAY28650 standard; Protein; 85 AA.
 AC AAY28650;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Murine chisel (Csl) protein.
 XX
 KW Chisel protein; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; murine;
 KW X chromosome; regulation; adaptive process; muscle homeostasis; nucleus;
 KW detection; diagnosis; propylaxis; treatment; skeletal myopathy;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KW differentiation; gene therapy; transgenic animal; drug screening;
 KW scallop regulatory myosin light chain.
 XX
 OS Mus sp.
 XX
 PN WO9950410-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 26-MAR-1999; 99WO-AU00220.
 XX

PR 27-MAR-1998; 98AU-0002634.
 XX
 PA (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GCHO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 PI Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX WPI; 1999-610852/52.
 DR N-PSDB; AAX90903.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 2; Fig 3; 157pp; English.
 XX
 CC The present sequence is the murine chisel protein (Csl), that is a
 CC member of the EF-Hand protein super family and is involved in signalling
 CC pathways. Csl protein is localized to the nucleus and does not show
 CC significant homology to any known protein. Structural homology between
 CC Csl and scallop regulatory myosin light chain is however detected. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 SQ Sequence 85 AA;
 Query Match 67.3%; Score 262.5; DB 20; Length 85;
 Best Local Similarity 70.3%; Pred. No. 5.5e-21;
 Matches 52; Conservative 5; Mismatches 12; Indels 5; Gaps 1;
 QY 1 MSKQPSNIRSIQANINIPMGAFRPGAGOPPRKKEFSTEEBOHVPTPSESEKSEKKRIP 60
 DB 1 MSKQPSNVRAIQANINIPMGAFRPGAGOPPRKKESTPTEBGAFT-----TSEKKRIP 55
 QY 61 GAVKLRGPAPNLSE 74
 DB 56 GMKKRPGPVNLSE 69
 RESULT 5
 ID ABG11263 standard; Protein; 172 AA.
 AC ABG11263;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #11254.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX

```

PA (HYSE-) HYSEQ INC.
XX
XX Dirmacac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAST75450.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 41622; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 172 AA;
SQ
Query Match 20.8%; Score 81; DB 22; Length 172;
Best Local Similarity 28.8%; Pred. No. 0.49;
Matches 19; Conservative 10; Mismatches 23; Indels 14; Gaps 1;
QY 6 ASNIRSIQANINIPMGAFRPGAGQ-----PPKRKESTEEQHVPTPSSE 51
DB 21 AQIVHAIVCGILMPISGFPPWVGQVFLALIQPRQSKTPSKKKEEEEEEEEEEE 80
QY 52 KSEEEK 57
DB 81 EEEEE 86
RESULT 6
ABG23750
ABG23750 standard; Protein; 602 AA.
XX
XX ABG23750;
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #23741.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX

```

[illegible]

XX OS Homo sapiens.
XX PN WC200155318-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001MO-US01332.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0180664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0186350.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229517.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244647.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0255719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.
XX

ID AAG49497 standard; Protein; 222 AA.
XX AAG49497;
AC
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 62625.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135114.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-014760.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148655.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0148426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.

```

PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 19.4%; Score 75.5; DB 21; Length 222;
 Best Local Similarity 28.4%; Pred. No. 2.6; Mismatches 21; Indels 19; Gaps 2;
 Matches 21; Conservative 13;

```

QY 1 MSKQASNIRSIQANINIPMGAFRPGAGQPPRKE-----FSTEEBOH 43
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 48 LSPAPAPGL--AQANGRLONGSFGSGSMIPQTKESWPESSSTTTDEEFKMATFDDEKN 105
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 44 VPTPESEKSEK 57
   |||||:|||||:|||||:|||||:|||||:|||||:
DB 106 TXLPAPAFEEERSE 119

```

RESULT 10
 ABB68168 standard; Protein; 1469 AA.
 ID ABB68168
 XX ABB68168;
 AC ABB68168;
 OS

```

XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 31296.
XX Drosophila melanogaster polypeptide SEQ ID NO 31296.
DE Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
KW Drosophila melanogaster.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PMD, Myers EW;
PI N-PSDB; ABL12271.
XX WPI; 2001-656860/75.
XX N-PSDB; ABL12271.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 31296; 21bp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS7072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1469 AA.
SQ

```

Query Match 19.4%; Score 75.5; DB 22; Length 1469;
 Best Local Similarity 32.3%; Pred. No. 21; Mismatches 21; Indels 5; Gaps 3;
 Matches 20; Conservative 16;

```

QY 3 KQPSNIRSIQANINIPMGAFRPGAGQPPRKE-----FSTEEBOHVPPESEKSEK 59
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1044 QKPTSPVAKLMPNINIVHLLPGSGVFKLRKQSSSSR--DBPQATVQTEAPPS 1101
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 PG 61
DB 1102 SG 1103

```

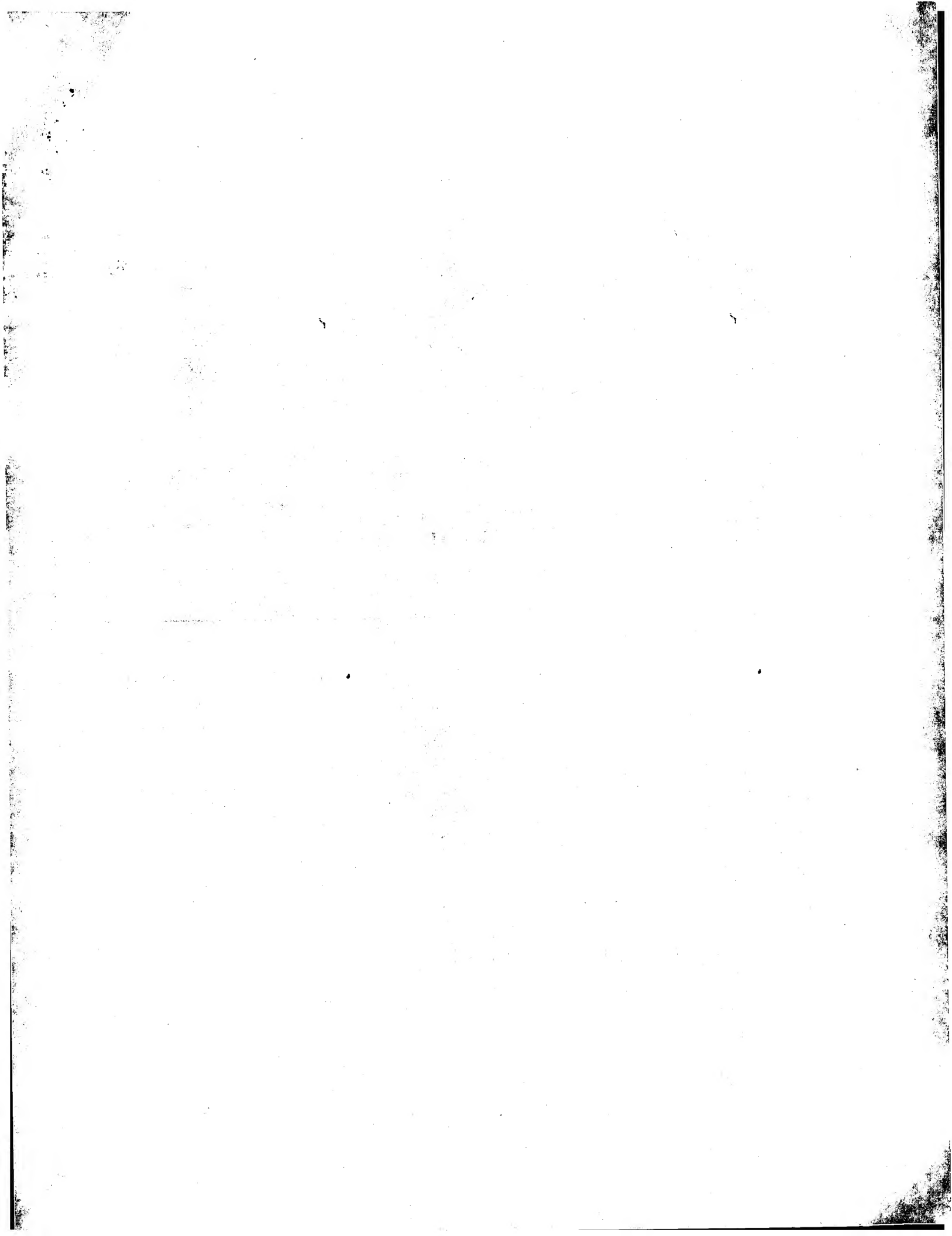
RESULT 11
 ABB25362 standard; Protein; 167 AA.
 ID ABB25362
 XX ABB25362;
 AC ABB25362;
 XX 18-FEB-2002 (first entry)
 XX Novel human diagnostic protein #25353.
 XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS

XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX N-PSDB; AAS89549.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 55721; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridization probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating
 XX disorders involving aberrant protein expression or biological activity.
 XX The polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. ABG00010-ABG030377 represent novel human
 XX diagnostic amino acid sequences of the invention.
 XX Note: The sequence data for this patent did not appear in the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 SQ Sequence 167 AA;
 Query Match 19.1%; Score 74.5; DB 22; Length 167;
 Best Local Similarity 38.1%; Pred. No. 2.4;
 Matches 24; Conservative 7; Mismatches 25; Indels 7; Gaps 3;
 QY 20 MGARPGAGGAPPRKE-FSTEEQHV--PTPESEKES-----EKKKPIFGAVKLPDPANL 72
 Db 91 MSSRPFGGNDSDRRVYTNHEWELDPDKDLSEIVQBEKKKLPEGNKVSORLKNL 150
 QY 73 SET 75
 Db 151 PHT 153
 RESULT 12
 AAR9451
 ID AAR9451 standard; protein; 244 AA.
 XX
 AC AAR9451;
 XX
 DT 17-FEB-1997 (first entry)
 XX
 DE Thymidine phosphorylase.
 XX
 XX Thymidine phosphorylase; endothelial cell growth factor; PD-ECGF;
 KW human platelet; vascular tissue; monoclonal antibody; Mab; immunoassay;

KW malignant tumour; metastasis; rheumatoid arthritis; diabetic cataract;
 KW premature cataract; senile macular degeneration.
 XX
 OS Homo sapiens.
 XX
 PN WO9620217-A1.
 XX
 PD 04-JUL-1996.
 XX
 PF 25-DEC-1995; 95WO-JP02661.
 XX
 PR 28-DEC-1994; 94JP-0327328.
 XX
 PA (TAIH) TAIHO PHARM CO LTD.
 XX
 PI Akiyama S, Miyadera K, Takebayashi Y, Yamada Y;
 XX WPI; 1996-321807/32.
 DR
 XX
 PT Monoclonal antibody against PD-ECGF derived peptide(s) - used in
 PT immunoassay of human platelet derived-PD-ECGF, for treatment of
 PT tumours etc.
 PT
 PS Claim 1; Page 14-15; 27pp; Japanese.
 XX
 XX The sequences given in AAR9451-52 represent thymidine phosphorylase
 XX and endothelial cell growth factor, respectively. These sequences
 XX are derived from human platelets and vascular tissue. A monoclonal
 XX antibody against these two sequences may be used in an immunoassay.
 XX The antibody may be used in the diagnosis and treatment of malignant
 XX tumours, metastasis, rheumatoid arthritis, diabetic cataract,
 XX premature cataract, and senile macular degeneration.
 XX
 SQ Sequence 244 AA;
 Query Match 19.1%; Score 74.5; DB 17; Length 244;
 Best Local Similarity 36.0%; Pred. No. 3.6;
 Matches 18; Conservative 6; Mismatches 21; Indels 5; Gaps 1;
 QY 25 PGAGPPRRKEFSTEEQHVPTPESEKESREKPIFGAVKLPDPANLSE 74
 Db 1 PGTAAPPAPDFSGSGQGLPDP-----SPEPKQLBELIMKRDGRLSE 45
 RESULT 13
 AAU02919
 ID AAU02919 standard; Protein; 323 AA.
 XX
 AC AAU02919;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Angiotensin converting enzyme (ACEV) splice variant protein #19.
 XX
 XX Angiotensin converting enzyme splice variant; ACEV; interleukin 6;
 KW granulocyte colony stimulating factor receptor; glucagon; hyperotropy;
 KW platelet-derived endothelial cell growth factor; cardiovascular disease;
 KW cellular tumour antigen P53; cyclin-dependent kinase inhibitor 1C;
 KW vasoactive intestinal polypeptide receptor 2; arteriosclerosis; cancer;
 KW myocardial infarction; coronary arterial thrombosis; renal disease;
 KW diabetic nephropathy; muscular disease; immune disorder; sarcoidosis;
 KW multiple sclerosis; immune complex nephritis; deep vein thrombosis;
 KW noncardiogenic pulmonary granulomatous disease; endothelial abnormality;
 KW vascular disorder; asbestosis.
 XX
 OS Homo sapiens.
 XX
 PN WO200136632-A2.
 XX
 PD 25-MAY-2001.
 XX
 DE 17-NOV-2000; 2000WO-IL00766.

KM nonaroidotic pulmonary granulomatous disease; endothelial abnormality;
 KM vascular disorder; asbestosis.
 OS Homo sapiens.
 PN WO200136632-A2.
 XX
 PD 25-MAY-2001.
 XX
 PF 17-NOV-2000; 2000MO-IL00766.
 XX
 PR 17-NOV-1999; 99IL-0132978.
 PR 10-DEC-1999; 99IL-0133455.
 XX
 PA (COMP-) COMPUGEN LTD.
 XX
 PI Levine Z, David A, Azar I, Khosravi R, Bernstein J;
 XX
 DR WPI; 2001-336004/35.
 DR N-PSDB; AAS06017.
 XX
 PT Novel alternative splicing variants e.g. variant of angiotensin
 PT converting enzyme (ACEV), useful in identifying candidate compounds
 PT capable of binding to the variant and to detect anti-variant antibodies
 XX
 PS Claim 4; Fig 17; 51pp; English.
 XX
 CC The sequence represents an angiotensin converting enzyme splice variant
 CC (ACEV) polypeptide. The polypeptides of the invention include variants of
 CC granulocyte colony stimulating factor, receptor, glucagon, interleukin 6,
 CC platelet-derived endothelial cell growth factor, cyclin-dependent kinase
 CC inhibitor 1C, cellular tumour antigen p53, and vasoactive intestinal
 CC polypeptide receptor 2. The polypeptides and their associated nucleic
 CC acids are useful for identification of variant sequences and detection of
 CC candidate compounds capable of binding the molecules. The sequences of
 CC the invention can be used in the treatment and diagnosis of various
 CC disorders including cardiovascular diseases such as arteriosclerosis,
 CC myocardial infarction and coronary arterial thrombosis, renal diseases
 CC such as diabetic nephropathy, muscular diseases such as hypertrophy,
 CC immune disorders such as immune complex nephritis, multiple sclerosis,
 CC cancer, sarcoidosis, nonaroidotic pulmonary granulomatous diseases such
 CC as asbestosis and vascular pathologies involving an endothelial
 CC abnormality such as deep vein thrombosis.
 XX
 SQ Sequence 363 AA;
 Query Match 19.1%; Score 74.5; DB 22; Length 363;
 Best Local Similarity 36.0%; Pred. No. 5.6;
 Matches 18; Conservative 6; Mismatches 21; Indels 5; Gaps 1;
 Oy 25 PGAGQPPRKRFSTBEQHPTPESESEKSEKKPIGAVKLGPAFNLSE 74
 Db 7 PGTGAPFAPGDFSGSGGLPDP-----SPRPKQLPELIRMKRDKGRNSE 51

Search completed: April 9, 2003, 12:35:33
 Job time : 28.9153 secs



Db 3 MSKQPVSNVRAIQANINIPMGAFRPGAGQPPRRKCTPEVEEYVP-PTSD---EKKKPIIP 58
QY 61 GAVVLPGPAPNLSE 74
|||
Db 59 GAKTLPGPAPNLSE 72

RESULT 2
POLY_EEEV STANDARD; PRT; 1239 AA.
AC P08768;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Structural polypeptide (P130) [Contains: Coat protein C (EC 3.4.21.-)
DE (Capitid protein C); Spike glycoprotein E3; Spike glycoprotein E2;
DE 6 kDa peptide; Spike glycoprotein E1].
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=82V-2137;
RX MEDLINE=87282265; PubMed=2886548;
RA Chang G.-J., Trent D.W.;
RT "Nucleotide sequence of the genome region encoding the 26S mRNA of
RT eastern equine encephalomyelitis virus and the deduced amino acid
RT sequence of the viral structural proteins.";
RL J. Gen. Virol. 68:2129-2142(1987).
CC -1- FUNCTION: THE CAPSID PROTEIN IS AN AUTO-PROTEASE.
CC -1- PFM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD NATURE PROTEINS.
CC -1- MISCELLANEOUS: THE 6 KDa POLYPEPTIDE PROBABLY SERVES AS THE SIGNAL
CC SEQUENCE FOR THE MEMBRANE GLYCOPROTEIN E1, WHICH IS THE VIRAL
CC HEMAGGLUTININ.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X05816; CAA29261.1; -.
CC PIR: A26816; VHWVER.
CC HSSP: P03315; IYCP.
DR MEROPS: S03.001; -.
DR InterPro: IPR002548; Alpha_E1_glycop.
DR InterPro: IPR000936; Alpha_E2_glycop.
DR InterPro: IPR002533; Alpha_E3_glycop.
DR InterPro: IPR001836; Alpha core.
DR InterPro: IPR000930; Togavirin.
DR Pfam: PF00943; Alpha_E2_glycop; 1.
DR Pfam: PF00944; Alpha_core; 1.
DR Pfam: PF01563; Alpha_E3_glycop; 1.
DR Pfam: PF01589; Alpha_E1_glycop; 1.
DR PRINTS: PR00798; TOGAVIRIN.
KW Coat protein; Polypeptide; Transmembrane; Glycoprotein; Hydrolase;
KW Serine protease.
FT CHAIN 1 259 COAT PROTEIN C.
FT CHAIN 260 322 SPIKE GLYCOPROTEIN E3.
FT CHAIN 323 742 SPIKE GLYCOPROTEIN E2.
FT CHAIN 743 798 6 KDa PEPTIDE.
FT CHAIN 799 1239 SPIKE GLYCOPROTEIN E1.
FT ACT_SITE 136 136 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT TRANSMEM 261 277 POTENTIAL.
FT TRANSMEM 684 701 POTENTIAL.
FT TRANSMEM 727 737 POTENTIAL.

FT TRANSMEM 777 798 POTENTIAL.
FT TRANSMEM 1211 1235 POTENTIAL.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 932 932 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1239 AA; 137431 MW; 8C7664AA05D2D41C CRC64;
Query Match 19.6%; Score 76.5; DB 1; Length 1239;
Best Local Similarity 39.3%; Pred. No. 8.5;
Matches 22; Conservative 9; Mismatches 18; Indels 7; Gaps 3;
QY 10 RSIOANINIPMGAFRPGAGQPPRRKCTPEVEEYVP-PTSD---EKKKPIIP 61
|||
Db 45 RSL-ANTILKORAPNPAGPAPKAPKAPSLSTETKKRPPPA--KKQKRPKPG 97

RESULT 3
MYPH_CHICK STANDARD; PRT; 537 AA.
AC Q05623;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-binding protein H (MyBP-H) (H-protein) (86 kDa protein).
GN MYBP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=93155224; PubMed=7679114;
RA Vaughan K.T., Weber F.E., Einheber S., Fischman D.A.;
RT "Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa
RT protein) reveals extensive homology with MyBP-C (C-protein) with
RT conserved immunoglobulin C2 and fibronectin type III motifs.";
RL J. Biol. Chem. 268:3670-3676(1993).
CC -1- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH
CC THICK MYOFILAMENTS IN THE A-BAND.
CC -1- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN
CC THE SLOW TONIC ALD MUSCLE. NOT DETECTED IN GIZZARD OR HEART.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. MYBP
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L05605; AAA21418.1; -.
CC HSSP: P13726; ZHFT.
DR InterPro: IPR000282; Gytoc_receptor_2.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR003962; FNIII repeat.
DR InterPro: IPR003066; Ig MHC.
DR InterPro: IPR003598; Ig C2.
DR InterPro: IPR003600; Ig_Like.
DR Pfam: PF00047; Ig; 2.
DR Pfam: PF00041; Ig; 2.
DR PRINTS: PR00014; FNTYPEIII.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00410; Ig_Like; 1.
DR SMART: SM00408; IgG2; 1.
KW Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;
KW Repeat.
FT DOMAIN 135 221 FIBRONECTIN TYPE-III 1.


```

FT  VARIANT      222      222      /FTid=VAR_007644.
FT  VARIANT      289      289      /FTid=VAR_007645.
FT  VARIANT      289      289      E -> A (IN NMNGB1).
FT  VARIANT      397      398      /FTid=VAR_007646.
FT  VARIANT      398      398      MISSING (IN NMNGB1).
FT  VARIANT      471      471      /FTid=VAR_007647.
FT  VARIANT      471      471      L -> S.
SQ  SEQUENCE      482 AA; 49981 MW; 0652FA132C3BDE28 CRC64;

Query Match      19.1%; Score 74.5; DB 1; Length 482;
Best Local Similarity 36.0%; Pred. No. 4.8;
Matches 18; Conservative 6; Mismatches 21; Indels 5; Gaps 1;

QY  25 PGAGQPPKREKFEETEEQHVPTPESEKSEKRPVPAVKLPAPFLSE 74
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db  7 PGTGAPAPGDFSGSGSGCLPDP-----SPPEKQLPFLIRKRGGRUSE 51

RESULT 5
ID  ABFL_MOUSE      STANDARD;      PRT; 3726 AA.
AC  061329;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE  (AT-binding transcription factor 1).
GN  ABFL.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BALB/MK X ICR; TISSUE=Brain;
RX  MEDLINE=96194902; PubMed=8654949;
RA  Ido A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,
RA  Hashimoto T., Morinaga T., Nishi S., Tamaoki T.;
RT  "Cloning of the CDNA encoding the mouse ABFL1 transcription factor.";
RL  Gene 168;227-231(1996).
CC  -1- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC  sequence of the enhancer element of the AFP gene.
CC  -1- SUBCELLULAR LOCATION: Nuclear.
CC  -1- SIMILARITY: CONTAINS 4 HOMEBOX DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL; D26046; BA05046.1; -
DR  HSSP; P20263; IOCp.
DR  TRANSFAC; T03881; -
DR  MGI; MGI:99948; Atbfl.
DR  InterPro; IPR001356; Homeobox.
DR  InterPro; IPR000822; Znf_C2H2.
DR  InterPro; IPR003604; Znf_U1.
DR  Pfam; PF00046; homeobox; 4.
DR  Pfam; PF00096; Zf-C2H2; 20.
DR  PRODOM; PD000010; Homeobox; 4.
DR  SMART; SM00389; HOX; 4.
DR  SMART; SM00355; ZNF_C2H2; 22.
DR  SMART; SM00451; ZNF_U1; 7.
DR  PROSITE; PS00027; HOMEBOX_1; 2.
DR  PROSITE; PS00071; HOMEBOX_2; 4.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR  PROSITE; PS0157; ZINC_FINGER_C2H2_2; 9.
KW  Transcription regulation; Activator; Zinc-finger; Metal-binding;

```

```

KW  DNA-binding; Homeobox; Nuclear protein; Repeat.
FT  ZN_FING      79      103      C2H2-TYPE.
FT  ZN_FING      282      305      C2H2-TYPE.
FT  ZN_FING      641      664      C2H2-TYPE.
FT  ZN_FING      672      695      C2H2-TYPE.
FT  ZN_FING      727      751      C2H2-TYPE.
FT  ZN_FING      805      829      C2H2-TYPE.
FT  ZN_FING      946      969      C2H2-TYPE.
FT  ZN_FING      985      1009      C2H2-TYPE.
FT  ZN_FING      1041      1065      C2H2-TYPE.
FT  ZN_FING      1089      1113      C2H2-TYPE.
FT  ZN_FING      1233      1256      C2H2-TYPE.
FT  ZN_FING      1262      1285      C2H2-TYPE.
FT  ZN_FING      1370      1395      C2H2-TYPE.
FT  ZN_FING      1411      1433      C2H2-TYPE.
FT  ZN_FING      1439      1462      C2H2-TYPE.
FT  ZN_FING      1555      1579      C2H2-TYPE.
FT  ZN_FING      1606      1630      C2H2-TYPE.
FT  ZN_FING      1990      2013      C2H2-TYPE.
FT  ZN_FING      2152      2211      HOMEBOX 1.
FT  ZN_FING      2249      2308      HOMEBOX 2.
FT  ZN_FING      2335      2358      C2H2-TYPE.
FT  ZN_FING      2539      2561      C2H2-TYPE.
FT  ZN_FING      2650      2709      HOMEBOX 3.
FT  ZN_FING      2720      2743      HOMEBOX 4.
FT  ZN_FING      2952      3011      HOMEBOX 4.
FT  ZN_FING      3032      3056      C2H2-TYPE.
FT  ZN_FING      3552      3576      C2H2-TYPE.
FT  ZN_FING      461      491      POLY-GLU.
FT  ZN_FING      771      785      POLY-ALA.
FT  ZN_FING      1314      1317      POLY-GLN.
FT  ZN_FING      1734      1748      POLY-GLN.
FT  ZN_FING      1794      1799      POLY-GLN.
FT  ZN_FING      1856      1863      POLY-GLN.
FT  ZN_FING      2044      2059      POLY-PRO.
FT  ZN_FING      2405      2408      POLY-ALA.
FT  ZN_FING      3216      3220      POLY-PRO.
FT  ZN_FING      3380      3409      POLY-GLN.
FT  ZN_FING      3412      3420      POLY-GLN.
FT  ZN_FING      3534      3550      POLY-GLY.
FT  ZN_FING      3620      3623      POLY-PRO.
FT  ZN_FING      3659      3662      POLY-SER.
SQ  SEQUENCE      3726 AA; 406567 MW; 915ACBE58A72C98 CRC64;

Query Match      19.0%; Score 74; DB 1; Length 3726;
Best Local Similarity 30.9%; Pred. No. 47;
Matches 17; Conservative 8; Mismatches 20; Indels 10; Gaps 1;

QY  24 RPAGQPPKREKFEETEEQHVPTPESEKSEKRPVPAVKLPAPFLSE 68
    ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db  2438 KFKQADPPSAQPNQOTQKQGFPEMQQLEQLKTNAPQPLPAPAPSLPQP 2492

RESULT 6
BAT2_HUMAN
ID  BAT2_HUMAN      STANDARD;      PRT; 2142 AA.
AC  P48634;
DT  01-FEB-1996 (Rel. 33, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DT  01-FEB-1996 (Rel. 33, Last annotation update)
DE  Large proline-rich protein BAT2 (HLA-B-associated transcript 2).
GN  BAT2.
OS  Homo sapiens (human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=T-cell;
RC  MEDLINE=90192810; PubMed=2156268;
RA  Banerji J., Sands J., Strominger J.L., Spies T.;
RT  "A gene pair from the human major histocompatibility complex encodes
RT  large proline-rich proteins with multiple repeated motifs and a

```

RT single ubiqlutin-like domain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).
 RP SEQUENCE OF 1-1860 FROM N.A.
 RX MEDLINE=93272029; PubMed=8499947;
 RA Irie F.J.M., Bougueteloret L., Pileur S., Caterina D., Primas G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,
 RA Cohen D.;
 RT "dense Alu clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment.";
 RL Nat. Genet. 3:137-145(1993).
 CC - FUNCTION: UNKNOWN.
 CC - TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M3509; AAA35585.1; -;
 DR EMBL; M3518; AAA35586.1; -;
 DR EMBL; M3512; AAA35586.1; JOINED.
 DR EMBL; M35025; CAA78744.1; -;
 DR PIR; B35098; B35098.
 DR PIR; S36152; S36152.
 DR Gene; HGNC;13918; BART2.
 DR MIM; 142580; -;
 KM Repeat.
 FT DOMAIN 519 524
 FT 636 657
 FT 684 688
 FT 699 704
 FT 814 821
 FT 1340 1345
 FT 1398 1403
 FT 1436 1442
 FT 1982 1991
 FT 41 1795
 FT 41 95
 FT 98 154
 FT 281 337
 FT 1740 1795
 FT 337 549
 FT 337 418
 FT 476 549
 FT 1899 2089
 FT 1899 1948
 FT 1965 2014
 FT REPEAT 2040 2089
 FT REPEAT 57 57
 FT CONFLICT 109 109
 FT CONFLICT 414 414
 FT CONFLICT 532 532
 FT CONFLICT 682 682
 FT CONFLICT 730 730
 FT CONFLICT 750 750
 FT CONFLICT 834 834
 FT CONFLICT 1035 1035
 FT CONFLICT 1068 1068
 FT CONFLICT 1285 1285
 FT CONFLICT 1400 1400
 FT CONFLICT 1611 1611
 FT CONFLICT 1729 1729
 FT SEQUENCE 2142 AA; 227640 MW; 33DDFL6B9B52420A CRC64;
 Query Match 18.8%; Score 73.5; DB 1; Length 2142;
 Best Local Similarity 32.1%; Pred. No. 29;
 Matches 25; Conservative 10; Mismatches 34; Indels 9; Gaps 4;
 QY 3 KQASINRSTQAININIMGAFPGAGPPRKKEPSTEEBOHVPTESEBKS---EKKKPI 59

DB 875 KEETLAQTGEAGRKLP--ASRSGAGPPPPRRR-SRETETWGRPPSSRRGIPPEEGCAP 921
 QY 60 P---GAVKLPGEPAFNISE 74
 DB 932 PRRAGPIKKPPPTKYEE 949
 RESULT 7
 Y429 HUMAN STANDARD; PRT; 356 AA.
 AC 043312;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical protein KIAA0429.
 GN KIAA0429.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98116555; PubMed=9455477;
 RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VIII.
 RT 78 new cDNA clones from brain which code for large proteins in
 RT vitro.";
 RL DNA Res. 4:307-313(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB007889; BAA24859.1; -;
 DR InterPro; IPR003124; WH2.
 DR Pfam; PF02205; WH2; 1.
 DR SMART; SM00246; WH2; 1.
 KM Hypothetical protein.
 SO SEQUENCE 356 AA; 38083 MW; 66307631FA63DF57 CRC64;
 Query Match 18.2%; Score 71; DB 1; Length 356;
 Best Local Similarity 37.7%; Pred. No. 7.4;
 Matches 23; Conservative 9; Mismatches 17; Indels 12; Gaps 4;
 QY 7 SNRSTQAININIMGAFPGAGPPRKKEPSTEEBOHVPTESEBKS-EKKKPIGAVKLP 66
 DB 263 SSWMSCQASVNPPL-----PKP-----SIPEHRQAIPESEADQEREP-PSATVSP 310
 QY 67 G 67
 DB 311 G 311
 RESULT 8
 PLE1 HUMAN STANDARD; PRT; 4684 AA.
 AC Q15149; Q16640; Q15148;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HDL).
 GN PLECT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX TISSUE=Placenta;
RC MEDLINE=96210632; PubMed=8633055;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24)";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
RX MEDLINE=96312447; PubMed=8698233;
RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
RA Bullrich F., Burgess R.E., Amano S., Hudson D.L., Owarick K.,
RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
RA Uitto J.;
RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
RT cDNA cloning and genomic organization";
RL Genes Dev. 10:1724-1735(1996).
RN [3]
RP VARIANT MD-EBS 1003-GLN-ALA-1005 DEL.
RX MEDLINE=97049959; PubMed=8894687;
RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,
RA Hachisaka H., Nishikawa T., McLean W.H.I., Uitto J.;
RT "Homozygous deletion mutations in the plectin gene (PLEC1) in patients
RT with epidermolysis bullosa simplex associated with late-onset
RT muscular dystrophy";
RL Hum. Mol. Genet. 5:1539-1546(1996).
RN [4]
RP VARIANT MD-EBS LEU-429 INS.
RX MEDLINE=11090821; PubMed=11159198;
RA Bauer J.W., Rouan P., Kofler B., Reznicek G.A., Kornacker I.,
RA Muss W., Hamelner R., Klausegger A., Huber A., Pohl-Gubio G.,
RA Wiche G., Uitto J., Hentner H.;
RT "A compound heterozygous one amino-acid insertion/nonsense mutation in
RT the plectin gene causes epidermolysis bullosa simplex with plectin
RT deficiency";
RL Am. J. Pathol. 158:617-625(2001).
CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROFILS AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSES OR
CC HEMIDESMOSES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO
CC MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE
CC CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE
CC FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3, ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN
CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B, WHEREAS BOTH THE N-
CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC -1- PTM: PHOSPHORYLATED BY CDC2, REGULATES DISSOCIATION FROM
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC -1- DISEASE: DEFECTS IN PLECT1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA
CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE
CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL
CC OF THE HEMIDESMOSE AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 PLECTIN REPEATS.
CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC -1-
CC EMBL; U53204; AAB05427.1; -
DR EMBL; U53610; AAB05428.1; -.
DR EMBL; U63609; AAB05428.1; JOINED.
DR EMBL; X97053; CAAB5765.1; -.
DR HSSP; Q01082; 1BKR.
DR Genem; HGNC:9069; PLECT1.
DR MIM; 601282; -.
DR MIM; 226670; -.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001715; Calponin-like.
DR InterPro; IPR001101; Plectin_repeat.
DR InterPro; IPR005326; S10_Plectin.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00307; CH_2.
DR Pfam; PF00681; Plectin; 19.
DR Pfam; PF03501; S10_Plectin; 1.
DR SMART; SM00033; CH_2.
DR SMART; SM00250; PLEC; 33.
DR SMART; SM00150; SPEC; 5.
DR PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR PROSITE; PS50021; CH; 2.
DR Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
DR Phosphorylation; Alternative splicing; Epidermolysis bullosa;
KW Disease mutation.
FT DOMAIN 1 1470
FT 1471 2755 GLOBULAR 1.
FT DOMAIN 2756 4684 CENTRAL FIBROUS ROD DOMAIN.
FT DOMAIN 175 400 GLOBULAR 2.
FT DOMAIN 179 282 ACTIN-BINDING.
FT DOMAIN 295 397 CH 1.
FT REPEAT 645 710 SPECTRIN 1.
FT REPEAT 740 824 SPECTRIN 2.
FT REPEAT 837 930 SPECTRIN 3.
FT REPEAT 1315 1415 SPECTRIN 4.
FT REPEAT 1469 2756 COILED COIL.
FT DOMAIN 2826 2863 COILED COIL (POTENTIAL).
FT REPEAT 2864 2901 PLECTIN 1.
FT REPEAT 2902 2939 PLECTIN 2.
FT REPEAT 2940 2977 PLECTIN 3.
FT REPEAT 2981 3015 PLECTIN 4.
FT REPEAT 3116 3153 PLECTIN 5.
FT REPEAT 3154 3191 PLECTIN 6.
FT REPEAT 3192 3229 PLECTIN 7.
FT REPEAT 3230 3267 PLECTIN 8.
FT REPEAT 3268 3305 PLECTIN 9.
FT REPEAT 3306 3343 PLECTIN 10.
FT REPEAT 3345 3382 PLECTIN 11.
FT REPEAT 3383 3420 PLECTIN 12.
FT REPEAT 3421 3458 PLECTIN 13.
FT REPEAT 3459 3496 PLECTIN 14.
FT REPEAT 3497 3534 PLECTIN 15.
FT REPEAT 3535 3572 PLECTIN 16.
FT REPEAT 3573 3610 PLECTIN 17.
FT REPEAT 3611 3648 PLECTIN 18.
FT REPEAT 3649 3686 PLECTIN 19.
FT REPEAT 3687 3724 PLECTIN 20.
FT REPEAT 3725 3762 PLECTIN 21.
FT REPEAT 3763 3800 PLECTIN 22.
FT REPEAT 3801 3838 PLECTIN 23.
FT REPEAT 3839 3876 PLECTIN 24.
FT REPEAT 3877 3914 PLECTIN 25.
FT REPEAT 3915 3952 PLECTIN 26.
FT REPEAT 3953 3990 PLECTIN 27.
FT REPEAT 3991 4028 PLECTIN 28.
FT REPEAT 4029 4066 PLECTIN 29.
FT REPEAT 4067 4104 PLECTIN 30.
FT REPEAT 4105 4142 PLECTIN 31.
FT REPEAT 4143 4180 PLECTIN 32.
FT REPEAT 4181 4218 PLECTIN 33.
FT REPEAT 4219 4256 BINDING TO INTERMEDIATE FILAMENTS (BY
FT REPEAT 4257 4294 SIMILARITY).
FT REPEAT 4295 4332 4 X 4 AA TANDEN REPEATS OF G-S-R-X.
FT REPEAT 4333 4370 PHOSPHORYLATION (BY CDC2) (BY
FT REPEAT 4371 4408 SIMILARITY).
FT DOMAIN 4625 4640
FT MOD_RES 4539 4539

```

```

FT  VARSPLIC 1 174
FT  VARSPLIC 409 412
FT  VARSPLIC 429 429
FT  VARSPLIC 1003 1005
FT  VARSPLIC 185 185
FT  VARSPLIC 259 259
FT  VARSPLIC 550 550
FT  VARSPLIC 560 560
FT  VARSPLIC 706 706
FT  VARSPLIC 886 886
FT  VARSPLIC 1002 1002
FT  VARSPLIC 1309 1309
FT  VARSPLIC 1321 1321
FT  VARSPLIC 1334 1334
FT  VARSPLIC 1534 1534
FT  VARSPLIC 1662 1662
FT  VARSPLIC 1688 1688
FT  VARSPLIC 1767 1767
FT  VARSPLIC 1789 1789
FT  VARSPLIC 1910 1910
FT  VARSPLIC 2154 2154
FT  VARSPLIC 2160 2160
FT  VARSPLIC 2215 2215
FT  VARSPLIC 2244 2244
FT  VARSPLIC 3027 3027
FT  VARSPLIC 3310 3310
FT  VARSPLIC 3361 3361
FT  VARSPLIC 3408 3408
FT  VARSPLIC 3447 3447
FT  VARSPLIC 3531 3531
FT  VARSPLIC 3580 3580
FT  VARSPLIC 3589 3589
FT  VARSPLIC 3596 3596
FT  VARSPLIC 3616 3616
FT  VARSPLIC 3686 3686

MVAQMLMPDQLRAIYEVLFREGVAVAKKORPRSLPHYP
GVTLQVWRAMASLARGLVRETPAMCHFFMYLTNEGIALH
ROYLHPREIVASLQVRVRVAVMMVAPRTPHVQAOVGL
GSPKRGPLPTEBORLRYRKRELEVSEPTVPATRTTILA
RPEPEPAPAT --> MSEGDAEVRAYSVDVSGSSGSPED
TLFWNLKTKTORSRRSGGAGNSNSVLDPAERAVIRIA (IN
ISOFORM 2 AND ISOFORM 3).
MISSING (IN ISOFORM 3).
L --> LL (IN MD-EBS).
/FTId=VAR_011336.
MISSING (IN MD-EBS).
/FTId=VAR_011337.
S --> F (IN REF. 2).
N --> D (IN REF. 2).
H --> N (IN REF. 2).
I --> V (IN REF. 2).
Q --> R (IN REF. 2).
N --> Y (IN REF. 2).
V --> A (IN REF. 2).
V --> L (IN REF. 2).
L --> L (IN REF. 2).
V --> L (IN REF. 2).
I --> M (IN REF. 2).
A --> T (IN REF. 2).
WLC --> RLR (IN REF. 2).
Q --> E (IN REF. 2).
L --> A (IN REF. 2).
K --> R (IN REF. 2).
K --> N (IN REF. 2).
S --> R (IN REF. 2).
R --> Q (IN REF. 2).
S --> E (IN REF. 2).
E --> A (IN REF. 2).
F --> L (IN REF. 2).
F --> L (IN REF. 2).
S --> A (IN REF. 2).
G --> A (IN REF. 2).
R --> S (IN REF. 2).
K --> Q (IN REF. 2).
E --> Q (IN REF. 2).
N --> H (IN REF. 2).
V --> A (IN REF. 2).

Query Match 18.2%; Score 71; DB 1; Length 4684;
Best Local Similarity 32.4%; Pred. No. 1.1e+02;
Matches 22; Conservative 8; Mismatches 26; Indels 12; Gaps 3;

QY 5 PASNRSIQANINIPMGAFRRGAGQPPKREKFEPTSEOHVPTPESEKSEKPKIPGAVK 64
DB 109 PAARTPHVQA-VQGPIS-----PKRGPLPTEQRLYRKRELEVSEPTVPATRTQ 160
QY 65 -----LRGP 68
DB 161 RTLARPGP 168

RESULT 9
PEP - DROME STANDARD; PRT; 716 AA.
AC PA1073; Q9VVU3; Q9VVU4;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein on ecdysone puffs.
GN PEP OR CG6143.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Oregon-R;
RX MEDLINE=91138953; PubMed=1899840;
RT "A unique zinc finger protein is associated preferentially with
active ecdysone-responsive loci in Drosophila."
RL Gene Dev. 5:188-200(1991).
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
Abail J.F., Asparyani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodera C., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
McKusick G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,
Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"the genome sequence of Drosophila melanogaster.";
Science 287:2185-2195(2000).
-1- FUNCTION: MAY PLAY A ROLE IN THE PROCESS OF EARLY AND LATE GENE
ACTIVATION, OR POSSIBLY IN RNA PROCESSING, FOR A DEFINED SET OF
DEVELOPMENTALLY REGULATED LOCI.
-1- SUBCELLULAR LOCATION: NUCLEAR, WHERE IT IS ASSOCIATED WITH THE
ACTIVE ECDYSONE-REGULATED LOCI ON POLYTENE CHROMOSOMES, AND ON
SOME HEAT SHOCK-INDUCED PUFFS. ITS DISTRIBUTION PATTERN FOLLOWS
THE CHANGES OF PUFFING PATTERNS IN THE DEVELOPMENTAL PROGRAM, OR
FOLLOWING HEAT SHOCK.
-1- DEVELOPMENTAL STAGE: MATERNAL PEP GENE TRANSCRIPTS ARE PROBABLY
SUPPLIED TO THE EMBRYO AT EARLY GENE ACTIVATION OR EXPRESSION. THE
ABUNDANCE OF TRANSCRIPTS THEN DECREASES TO A LOWER, FAIRLY
CONSTANT LEVEL THEREAFTER.
-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-24 IS THE INITIATOR.
This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sdb.ch/announce/
or send an email to license@isb-sdb.ch).
EMBL; X56689; CAA40017.1; -
DR EMBL; AE003523; AAF49317.1; -
DR PIR; A38703; A38703.

```


Db 512 PGASGLPPKAPDKEGSEGHKGPCKGSDSKCKK--FGSGKCGCPA 556

RESULT 12

ID_NFH_MOUSE STANDARD; PRT; 1087 AA.
 AC P19246; Q61959; 16, Created)
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 GN NEFH OR NFH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314(1988).
 [2]
 SEQUENCE FROM N.A.
 RX MEDLINE=89089138; PubMed=3145094;
 RA Sheldman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-231(1988).
 [3]
 SEQUENCE FROM N.A.
 RA Carden M.J.;
 RC STRAIN=Swiss Webster; TISSUE=Brain;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
 CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL; M24496; AAA39813.1; -
 CC EMBL; M23349; AAA39813.1; JOINED.
 CC EMBL; M24494; AAA39813.1; JOINED.
 CC EMBL; M24495; AAA39813.1; JOINED.
 CC EMBL; M31513; AAA39809.1; ALI_FRAME.
 CC EMBL; Z31012; CAA83229.1; -
 CC PIR; J10368; QFMNH.
 CC PIR; A43778; A43778.
 CC MGD; MGI:97309; NFH.
 CC InterPro; IPR001664; IF.
 CC Pfam; PF00038; Filament; 1.

DR PROSITE; P500226; IF, 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KM Repeat.

FT	DOMAIN	1	97	HEAD.
FT	DOMAIN	98	408	ROD.
FT	DOMAIN	409	1087	TAIL.
FT	DOMAIN	436	517	GLU-RICH (ACIDIC).
FT	DOMAIN	519	886	50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
FT	DOMAIN	887	1087	GLU/LYS-RICH.
FT	DOMAIN	98	129	COIL 1A.
FT	DOMAIN	130	141	LINKER 1.
FT	DOMAIN	142	239	COIL 1B.
FT	DOMAIN	240	261	LINKER 12.
FT	DOMAIN	262	283	COIL 2A.
FT	DOMAIN	284	287	LINKER 2.
FT	DOMAIN	288	408	COIL 2B.
FT	CONFLICT	133	133	K -> QA (IN REF. 2 AND 3).
FT	CONFLICT	199	199	A -> AR (IN REF. 2 AND 3).
FT	CONFLICT	281	281	S -> T (IN REF. 2 AND 3).
FT	CONFLICT	492	492	L -> G (IN REF. 2 AND 3).
FT	CONFLICT	551	551	P -> BREAKSP (IN REF. 3).
FT	CONFLICT	689	712	MISSING (IN REF. 3).
FT	CONFLICT	714	714	G -> A (IN REF. 3).
FT	CONFLICT	814	814	V -> M (IN REF. 2 AND 3).
FT	CONFLICT	843	843	T -> N (IN REF. 2 AND 3).
SO	SEQUENCE	1087 AA;	116612 MM;	57BAC76A38ED1C9 CRC64;

Query Match 17.7%; Score 69; DB 1; Length 1087;
 Best Local Similarity 37.5%; Pred. No. 37;
 Matches 15; Conservative 7; Mismatches 14; Indels 4; Gaps 1;

Oy 27 AGQPPKKEPSTEEBOVPTPESESEKKEKPIGAVKLP 66
 Db 893 AKPEPKK---VEBEKLPFKTEAKSKDPAKPAKAP 928

RESULT 13

ID_H1BN_XENLA STANDARD; PRT; 589 AA.
 AC P06180;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Histone-binding protein H1/N2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 OC Xenopodidae; Xenopus.
 NCBI_TaxId=8355;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Oocyte;
 RX MEDLINE=87161764; PubMed=3549279;
 RA Kleinschmidt J.A., Dingwall C., Maier G., Franke W.W.;
 RT "Molecular characterization of a karyophilic, histone-binding
 RT protein: cDNA cloning, amino acid sequence and expression of nuclear
 RT protein H1/N2 of Xenopus laevis.";
 RL EMBL J. 5:5547-3552(1986).
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN NUCLEOSOME ASSEMBLY. IT
 CC IS BOUND TO H3 AND H4 IN THE ABSENCE OF DNA, BUT RELEASED FROM
 CC H3 AND H4 IN THE PRESENCE OF DNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: WITH RABBIT NUCLEAR AUTONANTIGENIC SPERM PROTEIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL; X04712; CAA28419.1; -

DR PIR; A25680; A25680.
 DR InterPro; IPR001440; TPR.
 DR Pfam; PF00515; TPR; 3.
 KW Nuclear protein; Chromosomal protein.
 FT INIT MET 0 0
 FT DOMAIN 107 118 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 289 325 ASF/GLU-RICH (ACIDIC).
 FT DOMAIN 530 536 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 SQ SEQUENCE 589 AA; 64897 MW; AC28927B4D4E245B CRC64;
 Query Match 17.6%; Score 68.5; DB 1; Length 589;
 Best Local Similarity 31.4%; Pred. No. 22;
 Matches 18; Conservative 15; Mismatches 21; Indels 3; Gaps 2;
 QY 2 SKOPASINRISQAININIPMGAFRPGAGOPPRKKESTEEQHVPTEPES-EEKSEKK 57
 DB 182 TEKUTDILKIDSASRDVPM--KSGKGFPESKDAETLVEKESKPEITLKEKSIETK 236
 RESULT 14
 ID GISE HUMAN STANDARD; PRT; 720 AA.
 AC G9NY23; Q9BRE0; Q9Y557; Q9UG29;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE G2 and S phase expressed protein 1 (B99 homolog).
 GN GISE1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=20432105; PubMed=10974554;
 RA Monte M., Collavin L., Lazarevic D., Urrera R., Dragani T.A.,
 RA Schneider C.,
 RT "Cloning, chromosome mapping and functional characterization of a
 RT human homologue of murine Gtse-1 (B99) gene.",
 RL Gene 254:229-236 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smit L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baggeley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrell W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha W.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mareshgh-Mohammadi M., Matthews L.H., McCann O.T.,
 RA Mcclay J., McLaren S., McMurtry A.A., Milne S.A., Mottimore B.J.,
 RA Oeili C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soederlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M., Wilmer T.E., Wilming L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kuch J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren O., Shaull S., Sloan D., Song L.,
 RA Wang C., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,

RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinde K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
 RA Schert P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saito S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Peyrard M., Kedra D.,
 RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkerson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tikhonov V., Wright H.;
 RA "The DNA sequence of human chromosome 22.",
 RL Nature 402:489-495 (1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RU Submitted (APR-2001) to the EMBL/GenBank/DBJ database.
 CC -1- FUNCTION: May be involved in p53-induced cell cycle arrest in G2/M
 CC phase by interfering with microtubule rearrangements that are
 CC required to enter mitosis. Overexpression delays G2/M phase
 CC progression.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Associated with microtubules.
 CC -1- DEVELOPMENTAL STAGE: Expressed in G2/M phase. Not detected in
 CC quiescent cells.
 CC -1- PTM: Phosphorylated in mitosis (By similarity).
 CC* This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF223408; AAP31459.1; -;
 DR EMBL; AL031588; CAB38413.1; -;
 DR EMBL; AL022325; CAB63079.1; -;
 DR EMBL; BC006325; AAH06325.1; -;
 KW Microtubules; Phosphorylation.
 FT DOMAIN 22 27 POLY-SER.
 FT CONFLICT 259 259 V -> I (IN REF. 1).
 FT CONFLICT 506 506 R -> W (IN REF. 2; CAB38415).
 SQ SEQUENCE 720 AA; 76614 MW; ACD91CCDD008A99C CRC64;
 Query Match 17.6%; Score 68.5; DB 1; Length 720;
 Best Local Similarity 31.4%; Pred. No. 27;
 Matches 16; Conservative 9; Mismatches 25; Indels 1; Gaps 1;
 QY 25 PGAGQPRKKESTEEQHVP-PESEKSEKKPIPGAVLPGAPFNLS 74
 DB 230 PGAAEKKEKIPASPRRTKIPAEKSHRDVLPDKPAGAVNPAGSHLQ 280
 RESULT 15
 ID TRX2 HUMAN STANDARD; PRT; 2715 AA.
 AC G9UM06; Q9UK25; Q95836; Q9Y669; Q9Y668; O15022; Q96GP2; Q96IP3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trithorax homolog 2 (Mixed lineage leukemia gene homolog 2 protein).
 GN TRX2 OR HRX2 OR ML2 OR ML4 OR KIA0304.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Agrard P.O., Valvatne H., Jeannoulin F., Adamson A.,
 RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
 RA Lamerdin J., Chambon P., Losson R., Stewart A., Aasland R.;

"Mammalian trithorax- and ASH1-like proteins: putative chromatin regulators which contain PHD fingers and SET domains." ; Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

[2]

SEQUENCE FROM N.A. (LONG ISOFORM).

RA Lamerzin J.E., McCreedy P.M., Adelson A.W., Burkhardt-Schultz K., RA Garcia J.E., Kyle A., Ramirez M., Stillwagen S., Gaines J., Dangman L., RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O., RA Carraro A.V.,

RT "Sequence analysis of a 1 Mb region in human 19q13.1." ; RT Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE OF 111-2715 FROM N.A. (LONG-ISOFORM).

RC TISSUE=Testis, and leukocyte;

RX MEDLINE=20105772; PubMed=1063508;

RA Huntman D.G., Chin S.-F., Mulieris M., Batley S.J., Collins V.P., Wiedemann L.M., Aparicio S., Caldas C.;

RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines." ; RT Oncogene 18:7975-7984(1999).

[4]

SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).

RC TISSUE=Brain;

RX MEDLINE=97349984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima J., Ohira M., Seki N., RA Miyajima T., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro." ; RT DNA Res. 4:141-150(1997).

[5]

SEQUENCE OF 1918-2715 FROM N.A.

RP TISSUE=Brain, and Skin;

RA Straussberg R.;

RN Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

[6]

PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).

RC TISSUE=Placenta, and Bone marrow;

RX MEDLINE=9931983; PubMed=10409430;

RA Fitzgerald K.T., Diaz M.O.;

RL "MLL2: A new mammalian member of the trx/MLL family of genes." ; RL Genomics 59:187-192(1999).

-1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.

-1- SUBCELLULAR LOCATION: Nuclear (by similarity).

-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A TRUNCATED FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS. ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL BLOOD LYMPHOCTES, AND PLACENTA.

-1- DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.

-1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.

-1- SIMILARITY: CONTAINS 1 BROWDOMAIN.

-1- SIMILARITY: CONTAINS 1 SET DOMAIN.

-1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.

-1- SIMILARITY: CONTAINS 1 CXK-TYPE ZINC FINGER.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.ebi.ac.uk/announcements>)

-- or send an email to license@ebi.ac.uk.

EMBL, AJ007041, CAB45385.1, -

EMBL, AD000671, - NOT ANNOTATED_CDS.

EMBL, AB002302, BAA20763.2, -

EMBL, AF186605, AAD56420.1, -

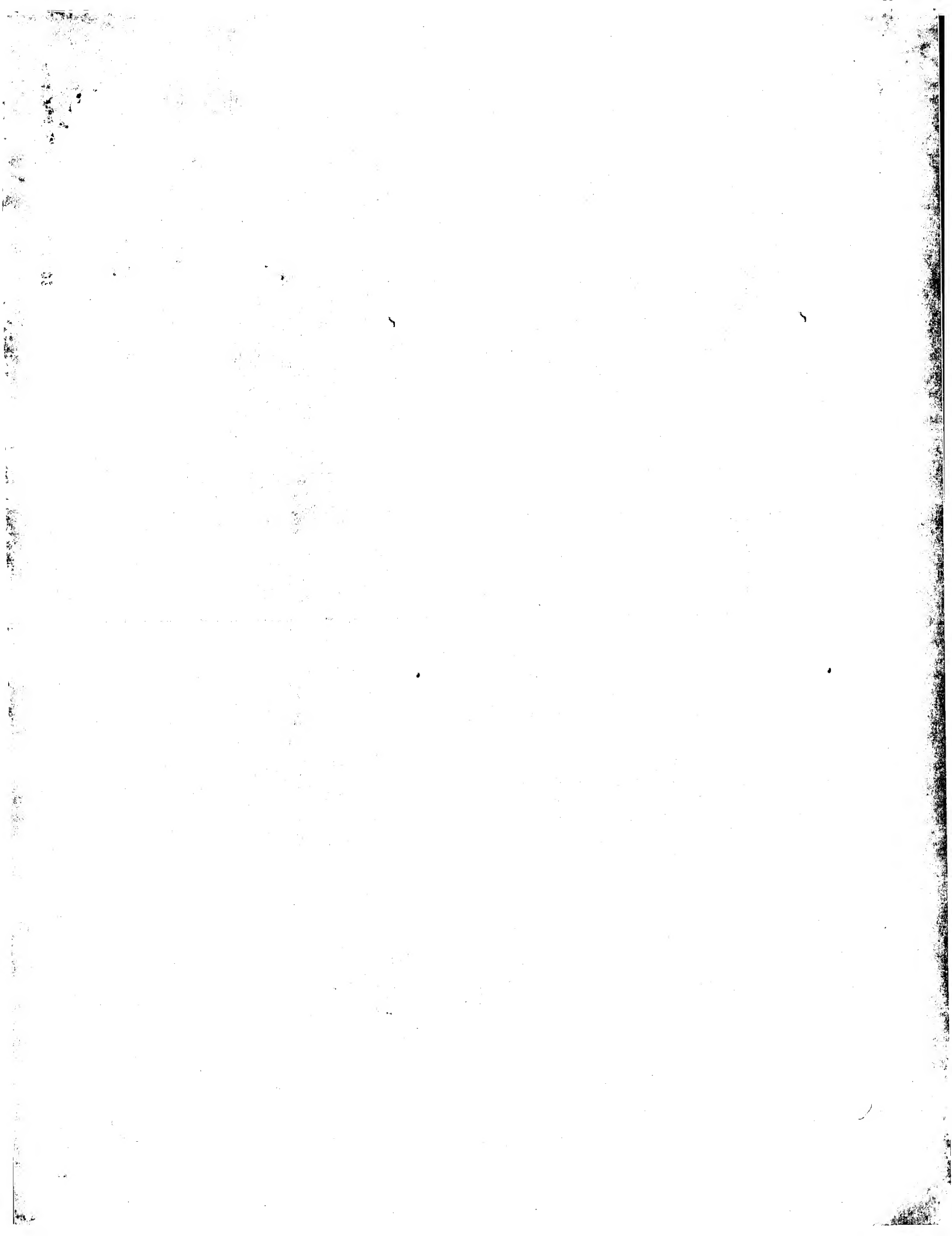
EMBL, AF104918, AAD17932.1, -

EMBL, AF105279, AAD26113.1, -

EMBL, BC009337, AAH09337.1, -

EMBL, BC007353, AAH07353.1, -

DB	1097	PGPPAPRRK	--IPRENELPLPEEEQSSRRKP	1127
DR	EMBL: AF105280; AAD26112.1; --			
DR	MTM; 606834; --			
DR	InterPro: IPR000637; AT_hook.			
DR	InterPro: IPR003889; FYrich_C.			
DR	InterPro: IPR003888; FYrich_N.			
DR	InterPro: IPR003616; PostSET.			
DR	InterPro: IPR001214; SET.			
DR	InterPro: IPR002857; ZnF_CXXC.			
DR	InterPro: IPR001965; ZnF_PHD.			
DR	InterPro: IPR001841; ZnF_ring.			
DR	Pfam: PF006628; SET; 3.			
DR	Pfam: PF00856; PHD; 1.			
DR	SMART; SM00208; zf-CXXC; 1.			
DR	SMART; SM00542; AT_hook; 1.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			
DR	PROSITE; PS01359; ZF_PHD_1; 3.			
DR	PROSITE; PS00516; ZF_PHD_2; 3.			
DR	SMART; SM00541; FYRN; 1.			
DR	SMART; SM00249; PHD; 4.			
DR	SMART; SM00508; PostSET; 1.			
DR	SMART; SM00184; RING; 1.			
DR	SMART; SM00317; SET; 1.			
DR	PROSITE; PS0280; SET; 1.			



GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using bw model

Run on: April 9, 2003, 12:31:56 ; Search time 38.4073 Seconds
(without alignments)
402.360 Million cell updates/sec

Title: US-09-647-019-5

Perfect score: 390
Sequence: 1 MSKOPASNIRSIOANINIPM.....KKRIPGAVKLPGPAPNLSET 75

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	385	98.7	91	13	Q90258
2	262.5	67.3	85	11	Q98R98
3	244.5	62.7	85	11	Q925F0
4	92.5	23.7	757	16	Q96QHO
5	80	20.5	602	10	Q9C5W9
6	80	20.5	613	10	Q9M355
7	79	20.3	2560	5	Q21980
8	77	19.7	574	5	017411
9	76.5	19.6	872	6	Q95LZ5
10	76.5	19.6	1321	16	Q86653
11	75.5	19.4	222	10	Q9ZWA4
12	75.5	19.4	480	13	Q91749
13	75.5	19.4	1469	5	Q9V953
14	75	19.2	231	11	Q99L83
15	75	19.2	759	11	Q8R184
16	74.5	19.1	482	4	Q8WVB7

17	73	18.7	660	11	Q9ET80	Q9ET80 mus musculus
18	72.5	18.6	1241	12	Q66579	Q66579 eastern equ
19	72	18.5	833	16	Q98063	Q98063 mycoplasma
20	5038	18.5	5038	11	Q90YX7	Q90YX7 mus musculus
21	71.5	18.3	307	16	Q8UIX8	Q8UIX8 agrobacteri
22	71	18.2	473	4	Q8TCA2	Q8TCA2 homo sapien
23	71	18.2	755	4	Q96RX2	Q96RX2 homo sapien
24	71	18.2	868	4	Q9UPU7	Q9UPU7 homo sapien
25	5085	18.2	5085	11	Q9JK56	Q9JK56 rattus norv
26	70.5	18.1	663	10	Q94HR8	Q94HR8 oryza sativ
27	70	17.9	227	6	Q97770	Q97770 oryctolagus
28	70	17.9	445	10	Q41733	Q41733 zea mays (m
29	69.5	17.8	316	2	Q50887	Q50887 myxococcus
30	69.5	17.8	609	17	Q8TXA4	Q8TXA4 methanopyru
31	69	17.7	414	2	Q9K334	Q9K334 salmonella
32	69	17.7	414	16	Q82113	Q82113 salmonella
33	69	17.7	447	10	Q8SAB2	Q8SAB2 sorghum bic
34	69	17.7	930	11	Q925G9	Q925G9 mus musculus
35	69	17.7	1557	5	Q8SY35	Q8SY35 drosophila
36	69	17.7	1674	5	Q8SY35	Q8SY35 drosophila
37	69	17.7	2837	3	Q96U29	Q96U29 neurospora
38	68.5	17.6	712	4	Q9Y557	Q9Y557 homo sapien
39	68.5	17.6	720	4	Q9NY23	Q9NY23 homo sapien
40	68.5	17.6	720	4	Q9BRE0	Q9BRE0 homo sapien
41	68.5	17.6	1060	3	Q9C259	Q9C259 neurospora
42	68.5	17.6	2664	5	Q26033	Q26033 plasmodium
43	68	17.4	415	5	Q960K6	Q960K6 drosophila
44	68	17.4	676	6	Q95UC9	Q95UC9 sus scrofa
45	68	17.4	682	5	Q9U010	Q9U010 drosophila

ALIGNMENTS

RESULT 1

ID	Q90258	PRELIMINARY;	PRT;	91 AA.
AC	Q90258;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	Chisel.			
GN	CSL.			
OS	Xenopus laevis (African clawed frog).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;			
OC	Xenopodidae; Xenopus.			
OX	NCBI_TaxID=8355;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21275706; PubMed=1181084;			
RA	Palmer S., Groves N., Schindeler A., Yeoh T., Biben C., Wang C.-C.,			
RA	Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,			
RA	Mohun T.J., Harvey R.P.;			
RT	"The Small Muscle-specific Protein Csl Modifies Cell Shape and			
RT	Promotes Myocyte Fusion in an Insulin-like Growth Factor 1-dependent			
RT	Manner";			
RL	J. Cell Biol. 153:985-998(2001).			
DR	EMBL, AF343894; AAK71068.1; -			
SO	SEQUENCE 91 AA; 1006 MW; BD2BA90B82B3846C CRC64;			

Query Match 98.7%; Score 385; DB 13; Length 91;

Best Local Similarity 100.0%; Pred. No. 8.1e-36; Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSKOPASNIRSIOANINIPMGAPFGAPGPPKXKESTEEHQVPTPESEKSEKKKPIP 60
DB	1	MSKOPASNIRSIOANINIPMGAPFGAPGPPKXKESTEEHQVPTPESEKSEKKKPIP 60
QY	61	GAVKLPAPNLSE 74
DB	61	GAVKLPAPNLSE 74

```

RESULT 2
OQ9ER98      PRELIMINARY;      PRT;      85 AA.
AC OQ9ER98;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Stretch responsive muscle (X-chromosome) (SMPX protein)
DE (muscle-specific protein CSL).
GN SMPX OR SRMX OR CSL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/10; TISSUE=SKLETTAL MUSCLE;
RX MEDLINE=21295047; PubMed=11401441;
RA Kemp T.J., Sadusky T.J., Simon M., Brown R., Eastwood R.,
RA Sassoon D.A., Coulton G.R.,
RT "Identification of a Novel Stretch-Responsive Skeletal Muscle Gene
RT (Smpx).";
RL Genomics 72:260-271(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065879; PubMed=10598820;
RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.,
RT "Identification, mapping, and genomic structure of a novel X-
RT chromosomal human gene (SMPX) encoding a small muscular protein.";
RL Hum. Genet. 105:506-512(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Patzak D.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=21275706; PubMed=11381084;
RA Palmer S., Groves N., Schindeler A., Yeoh T., Bilen C., Wang C.-C.,
RA Sparrow D.B., Barnett L., Jenkins N.A., Copeland N.G., Koentgen F.,
RA Mohun T., Harvey R.P.;
RT "The Small Muscle-Specific Protein Csl Modifies Cell Shape and
RT Promotes Myocyte Fusion in an Insulin-like Growth Factor-1-dependent
RT Manner";
RL J. Cell Biol. 153:985-998(2001).
DR EMBL; AJ245772; CAC08493.1; -.
DR EMBL; AF364070; AAK50398.1; -.
DR EMBL; AY026524; AAK07682.1; -.
DR MGI; MGI:1913356; Smpx.
SQ SEQUENCE 85 AA; 9253 MW; 43863840A65DA6BC CRC64;

Query Match 67.3%; Score 262.5; DB 11; Length 85;
Best Local Similarity 70.3%; Pred. No. 4e-22;
Matches 52; Conservative 5; Mismatches 12; Indels 5; Gaps 1;

QY 1 MSKOPASNISIOANINIPMGAFRPGAGOPPKRKEPSTEEBQHVP-TPSESEKSEKKPI 60
DB 1 MSKOPISNVASIOANINIPMGAFRPGAGOPPKRKEPSTEEBQAPT-----TSEKKKPI 55
QY 61 GAVKLPGAPFNLS 74
DB 56 GKKKPGPVVNLSE 69

RESULT 3
OQ25FO      PRELIMINARY;      PRT;      85 AA.
AC OQ25FO;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE SMPX protein.

```

```

GN SMPX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20065879; PubMed=10598820;
RA Patzak D., Zhuchenko O., Lee C.C., Wehnert M.;
RT "Identification, mapping, and genomic structure of a novel X-
RT chromosomal human gene (SMPX) encoding a small muscular protein.";
RL Hum. Genet. 105:506-512(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Patzak D.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF364071; AAK50399.1; -.
SQ SEQUENCE 85 AA; 9120 MW; 52654F8C790C932C CRC64;

Query Match 62.7%; Score 244.5; DB 11; Length 85;
Best Local Similarity 66.7%; Pred. No. 4.2e-20;
Matches 50; Conservative 5; Mismatches 13; Indels 7; Gaps 2;

QY 1 MSKOPASNISIOANINIPMGAFRPGAGOPPKRKEPSTEEBQHVP-TPSESEKSEKKPI 59
DB 1 MSKOPISNVASIOANINIPMGAFRPGAGOPPKRKEPSTGTABGAPATP-----EEKKPV 54
QY 60 GAVKLPGAPFNLS 74
DB 55 GKKKPGPVVNLSE 69

RESULT 4
OQ9EOHO      PRELIMINARY;      PRT;      757 AA.
AC OQ9EOHO;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein MYPV_3930.
GN MYPV 3930.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Bacillales; Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Champaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13566.1; -.
DR MYPULAC; MYPV_3930; -.
DR InterPro; IPR002965; P rich extensn.
DR PRINTS; PR01217; PRICHEXTENGN.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 757 AA; 86350 MW; BBD9FC7CD714238C CRC64;

Query Match 23.7%; Score 92.5; DB 16; Length 757;
Best Local Similarity 31.0%; Pred. No. 0.054;
Matches 26; Conservative 10; Mismatches 29; Indels 19; Gaps 2;

QY 5 PASNISIOANIN-----IPMGAFRPGAGOPPK-----RKFSSTEEBQHVP 45
DB 26 KXNLLYSEANISNTDFTNLEIPKPLPKPEPPKPPQEDPIIKIPQPEPKPEEPKPEP 85
QY 46 TPSESEKSEKKPIPGAVKLPGPA 69
DB 86 KPEPKKPEKPKAPPPVVIAPDA 109

```



```

RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RT Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N.;
RT "The sequence of C. elegans cosmid R13F6."
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U00046; AAC47044.3; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001258; NHL.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF01436; NHL; 5.
DR PRINTS; PR00011; EGFAMININ.
DR SMART; SM00181; EGF; 6.
DR PROSITE; PS00022; EGF_1; UNKNOWN_7.
DR PROSITE; PS01186; EGF_2; 7.
KM EGF-like domain; Glycoprotein; Hypohectical protein.
SQ SEQUENCE 2560 AA; 282782 MW; DDCB8ED4983CF4F CRC64;

Query Match 20.3%; Score 79; DB 5; Length 2560;
Best Local Similarity 35.7%; Pred. No. 6.9;
Matches 20; Conservative 8; Mismatches 24; Indels 4; Gaps 2;

OY 14 ANINIPMGAFPGAGPPKKEFSTEEQHPTPESEKSEK--KPIPAVKLP 66
Db 447 AYAVHPQAQSPRRQGETPSS-KTRKAQVKRPTSEKKEKSRLOKPIATVQVP 501

RESULT 8
017411 PRELIMINARY; PRT; 574 AA.
AC 017411;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DN Probable chitinase 1 precursor (EC 3.2.1.14).
GN CHTL.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Aedes.
OC NCBI_TaxID=7159;
OX RN
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98324849; PubMed=9662472;
RA de la Vega H., Specht C.A., Liu Y., Robbins P.W.;
RT "Chitins are a multi-gene family in Aedes, Anopheles and
RT Drosophila."
RL Insect Mol. Biol. 7:233-239 (1998).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC EMBL; AF026491; AAB81849.1; -.
DR InterPro; IPR001579; Chitinase_18/2.
DR InterPro; IPR002557; Chitin_bind_pera.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF01607; CBM_14; 1.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR SMART; SM00494; CHBD; 1.
DR PROSITE; PS01095; CHITINASE_18; 1.
KM Hydrolyase; Glycosidase; Chitin degradation; Glycoprotein;

```

```

KM Multigene family; Signal.
FT SIGNAL 1 15
FT CHAIN 16 574
FT DOMAIN 16 350
FT ACT SITE 175 175
FT CARBOHYD 83 83
FT CARBOHYD 301 301
SQ SEQUENCE 574 AA; 64268 MW; EALDD40AE9EFF0C CRC64;

Query Match 19.7%; Score 77; DB 5; Length 574;
Best Local Similarity 31.3%; Pred. No. 2.1;
Matches 26; Conservative 9; Mismatches 20; Indels 28; Gaps 5;

OY 4 OPASNISIQANINIPMGAFPGAGP-----PK-----RKPFSTEEQHPT 46
Db 413 RPPSTQTSIQ---EVPL-----AGGPTSTYTRRPKPTTAANKTRK--STTTTTPA 461

OY 47 PESEKSEKKKPIGAVKLP 69
Db 462 PDSSEEDRQPEPAPVPIPA 484

RESULT 9
095125 PRELIMINARY; PRT; 872 AA.
AC 095125;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypohectical 98.6 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopitheciinae; Macaca.
OC NCBI_TaxID=9541;
OX RN
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071046; BAB64439.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 7.
KM Hypohectical protein.
SQ SEQUENCE 872 AA; 98649 MW; 3DDDC4BE85480BDB CRC64;

Query Match 19.6%; Score 76.5; DB 6; Length 872;
Best Local Similarity 38.6%; Pred. No. 3.9;
Matches 27; Conservative 7; Mismatches 23; Indels 13; Gaps 5;

OY 12 IOANINIPMGAFPGAG---GPPKKEFSTEEQ-HVTPPESEKSEKKPIPAVKL- 65
Db 555 LQATPEIPQ--VPPGSGQAAPKKEEKEEQQKEEQQKEEKKP-----KLT 607

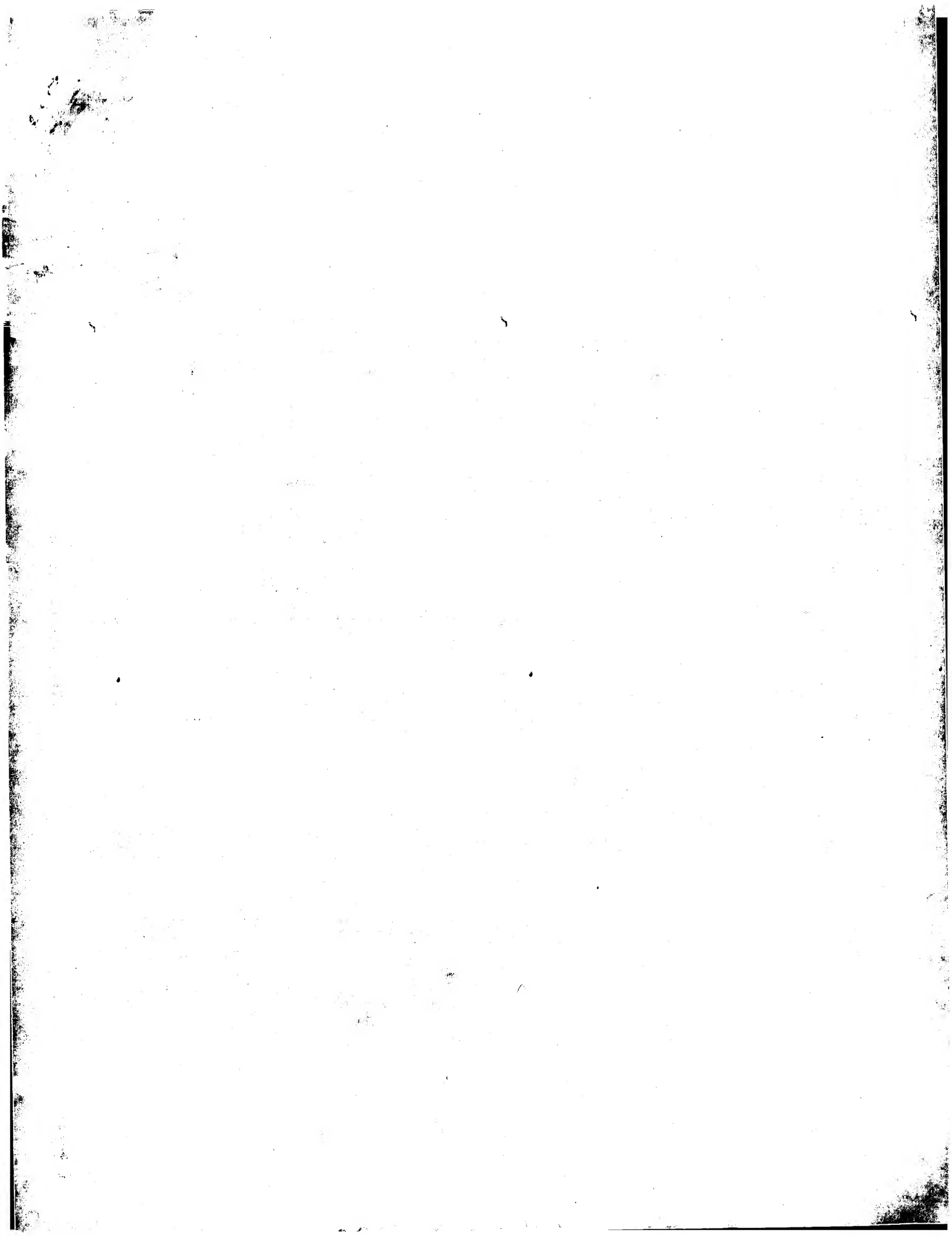
OY 66 PGPAFNLSET 75
Db 608 PSKVASLSES 617

RESULT 10
086653 PRELIMINARY; PRT; 1321 AA.
AC 086653;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATP/GTP binding protein.
GN SC05734 OR SC3C3.20C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
 RN NCB1_TaxID=1902;
 ON (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=A3(2);
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RL MEDLINE=97000351; PubMed=843436;
 RX Redenbach M., Kleiser H.M., Denepalte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble L., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2)."
 RL Nature 417:141-147(2002).
 DR EMBL; AL031231; CNA20269.1;
 DR InterPro; IPR002543; Pfam; PF0580; Pfam; PF0580; Pfam; PF0580;
 DR Pfam; PF0580; Pfam; PF0580;
 DR SEQUENCE 1321 AA; 143529 MW; 65C36645E4E7C6ED CRC64;
 Query Match 19.4%; Score 76.5; DB 16; Length 1221;
 Best Local Similarity 33.9%; Pred. No. 6.2;
 Matches 21; Conservative 8; Mismatches 26; Indels 7; Gaps 2;
 Oy 16 INIPGARGPGAGPPKKEPSTEEQHVPTESEKSEKPKPGAVK-----PGAPF 70
 Db 1013 LNVPTGV--PGRGVPRKLTHTALPRDSTPDSLSSEATQAQVQKVMAGPPAPFV 1070
 Oy 71 NL 72
 Db 1071 RL 1072
 RESULT 11
 Q9ZMA4 PRELIMINARY; PRT; 222 AA.
 AC Q9ZMA4;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DE 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE P11M21.25 protein.
 GN P11M21.25.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsi.
 ON NCB1_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Federlepel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altieri H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,

RA Gonzalez A., Kremenetskaya I., Kim C., Lenz C., Li J., Liu S.,
 RA Lueros S., Schwartz J., Shin P., Toriumi M., Vysotskaya V.S.,
 RA Walker M., Yu G., Becker J., Theologis A., Davis R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; ACC03027; M1010686.1;
 DR SEQUENCE 222 AA; 25205 MW; 4E8A390B611E468 CRC64;
 Query Match 19.4%; Score 75.5; DB 10; Length 222;
 Best Local Similarity 28.4%; Pred. No. 1.1;
 Matches 21; Conservative 13; Mismatches 21; Indels 19; Gaps 2;
 Oy 1 MSKPPANINISQANINIPWAGPPGAGPPKKE-----FSTEEQH 43
 Db 48 LSPAPAPGL--AQANGRLGNGSPGSGMIPTKESWSPSSSTTTDEFEKLMATPDEKN 105
 Oy 44 VPTSESEKSEKK 57
 Db 106 TKLPEAFEEEESE 119
 RESULT 12
 Q91749 PRELIMINARY; PRT; 480 AA.
 AC Q91749;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE Oviduct specific protein-1A.
 GN FOSP-1A.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 ON NCB1_TaxID=8355;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=OVIDUCT;
 RX MEDLINE=95262571; PubMed=7744034;
 RA Mehra R.J., Tata J.R.;
 RT "Structural and functional characterization and cloning of Xenopus
 FOSP-1 (frog oviduct-specific protein-1) gene."
 RL Eur. J. Biochem. 229:224-232(1995).
 DR EMBL; X81569; CAA57265.1;
 DR EMBL; X81571; CAA57267.1;
 DR SEQUENCE 480 AA; 53182 MW; 191204199FD13A65 CRC64;
 Query Match 19.4%; Score 75.5; DB 13; Length 480;
 Best Local Similarity 48.6%; Pred. No. 2.6;
 Matches 17; Conservative 6; Mismatches 9; Indels 3; Gaps 1;
 Oy 27 AGPPKKEPSTEEQHVPTESEKSEKPKPG 61
 Db 373 AAQVAKSESESEHVSSEKSEKPKER---PG 404
 RESULT 13
 Q9V953 PRELIMINARY; PRT; 1469 AA.
 AC Q9V953;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CG16742 protein.
 GN CG16742.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCB1_TaxID=7227;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;



GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 18:42:01 ; Search time 10.1771 Seconds

(without alignments)
5183.040 Million cell updates/sec

Title: US-09-647-019-6

Perfect score: 172

Sequence: 1 gggtcccaatcacggagagag.....taacaataatcaatgaagcag 172

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15333831 residues

Total number of hits satisfying chosen parameters: 682724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	86.6	909	US-09-484-970B-111	Sequence 111, App
2	30.2	17.6	444	US-08-464-342-1	Sequence 1, Appl
3	30.2	17.6	444	US-08-305-520-1	Sequence 1, Appl
4	30.2	17.6	444	US-08-464-604A-1	Sequence 1, Appl
5	30.2	17.6	444	US-08-875-272-1	Sequence 1, Appl
6	30.2	17.6	444	US-08-486-663A-1	Sequence 1, Appl
7	30.2	17.6	444	US-08-903-386-1	Sequence 1, Appl
8	30.2	17.6	444	US-08-247-904B-1	Sequence 1, Appl
9	30.2	17.6	444	US-08-895-601-4	Sequence 1, Appl
10	30.2	17.6	444	US-08-767-942A-1	Sequence 1, Appl
11	30.2	17.6	452	US-08-425-299A-4	Sequence 4, Appl
12	29.2	17.0	769	US-09-535-008-50	Sequence 50, Appl
13	28.8	16.7	2343	US-09-484-970B-119	Sequence 119, App
14	27	15.7	1480	US-08-839-008-8	Sequence 8, Appl
15	27	15.7	1537	US-08-839-008-1	Sequence 1, Appl
16	26.2	15.2	1432	US-09-183-861-73	Sequence 73, Appl
17	26.2	15.2	1432	US-09-022-765-73	Sequence 73, Appl
18	25.8	15.0	446	US-09-137-946-9	Sequence 9, Appl
19	25.8	15.0	446	US-08-680-506-9	Sequence 9, Appl
20	25.8	15.0	954	US-08-680-506-5	Sequence 5, Appl
21	25.8	15.0	1347	US-08-680-506-6	Sequence 8, Appl
22	25.8	15.0	1587	US-08-680-506-6	Sequence 8, Appl
23	25.8	15.0	2233	US-08-680-506-4	Sequence 4, Appl
24	25.8	15.0	5027	US-08-680-506-2	Sequence 2, Appl
25	25.6	14.9	1500	PCT-US93-08386-3	Sequence 3, Appl
26	25.2	14.7	388	US-07-626-618A-1	Sequence 1, Appl
27	25.2	14.7	388	US-07-928-611-1	Sequence 1, Appl

28	25.2	14.7	388	US-08-333-977-1	Sequence 1, Appl
29	25.2	14.7	388	US-08-487-811A-1	Sequence 1, Appl
30	25.2	14.7	388	US-09-060-694-1	Sequence 1, Appl
31	25.2	14.7	388	US-09-378-074-1	Sequence 1, Appl
32	25.2	14.7	388	PCT-US93-07370-1	Sequence 1, Appl
33	25.2	14.7	1161	US-08-086-439C-2	Sequence 2, Appl
34	25.2	14.7	1161	US-08-434-877-2	Sequence 2, Appl
35	25.2	14.7	1367	US-08-475-742-3	Sequence 3, Appl
36	25.2	14.7	1370	US-08-056-051-1	Sequence 1, Appl
37	25.2	14.7	1370	US-07-928-611-17	Sequence 17, Appl
38	25.2	14.7	1370	US-08-487-811A-17	Sequence 17, Appl
39	25.2	14.7	1370	US-09-060-694-17	Sequence 17, Appl
40	25.2	14.7	1370	US-09-378-074-17	Sequence 17, Appl
41	25.2	14.7	1370	PCT-US93-07370-17	Sequence 17, Appl
42	25.2	14.7	1466	US-08-056-051-3	Sequence 3, Appl
43	25.2	14.7	1466	US-07-928-611-19	Sequence 19, Appl
44	25.2	14.7	1466	US-08-487-811A-19	Sequence 19, Appl
45	25.2	14.7	1466	US-09-060-694-19	Sequence 19, Appl

ALIGNMENTS

```

RESULT 1
US-09-484-970B-111
; Sequence 111, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 111
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 021656.2CB1
US-09-484-970B-111

Query Match      86.6%; Score 149; DB 4; Length 909;
Best Local Similarity 98.8%; Pred. No. 7.4e-43;
Matches 171; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 2 GTTCTCAATACCGGAGAGGACACAGACTATTTCAGCCACATGAAAAAGCATCGGAATTGA 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36 GTTCTCAATACCGGAGAGGACACAGACTATTTCAGCCACATGAAAAAGCATCGGAATTGA 95

QY 62 GATCGACGCTCAAGAGACACC-GGGCGCCCTTCACACTTCGAGAGAGCTTTGATTCTT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 GATCGACGCTCAAGAGACACC-GGGCGCCCTTCACACTTCGAGAGAGCTTTGATTCTT 155

QY 121 GCATCT-GGCTGCTGGGACTTCCTTGGCAGTAAACAAATACATTAACAG 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 156 GCATCTGGGCTGCTGGGACTTCCTTGGCAGTAAACAAATACATTAACAG 208

RESULT 2
US-08-464-342-1/C
; Sequence 1, Application US/08464342
; Patent No. 5650313
; GENERAL INFORMATION:
; APPLICANT: NI, ET AL.
; TITLE OF INVENTION: Ubiquitin Conjugating Enzymes
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESSES:

```

Query Match	17.6%	Score 30.2;	DB 1;	Length 444;
Best Local Similarity	60.2%;	Pred. No. 0.25;		
Matches 50; Conservative	0;	Mismatches 33;	Indels 0;	Gaps 0;

[illegible]

RESULT 4
 US-08-464-604A-1/c
 : Sequence 1, Application US/08464604A
 : Patent No. 5849286
 :
 : GENERAL INFORMATION:
 :
 : * APPLICANT: NI, JIAN
 : APPLICANT: GENTZ, REINER
 : APPLICANT: ADAMS, MARK D
 : TITLE OF INVENTION: UBIQUITIN CONJUGATING ENZYMES 7, 8 AND 9
 : NUMBER OF SEQUENCES: 27
 : CORRESPONDENCE ADDRESSES:

STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,604A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PERRARO, GREGORY D
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-419
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STANDNESS: double

TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..441
US-08-464-604A-1

Query Match 17.6%; Score 30.2; DB 2; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

3 TTCTCAATACCGGAGAGGACAGAGCTATTTCACCCACATGAAAAGCATCGGAATTGAG 62
336 TTCTCAATCTGTTTGTATGATCCGAGCATCTCAGGACACTAAAGATCATCTGATTTGGG 337
63 ATCGAGCTCAGAGACACCGGG 85
336 ATCACACAACAGAGAACAGATGG 314

RESULT 5

US-08-875-272-1/c
Sequence 1, Application US/08875272
Patent No. 5945321
GENERAL INFORMATION:
APPLICANT: NI, ET AL.
TITLE OF INVENTION: Ubiqutin Conjugating Enzymes 7, 8 and 9
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESS: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,272
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-244
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-875-272-1

Query Match 17.6%; Score 30.2; DB 2; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

3 TTCTCAATACCGGAGAGGACAGAGCTATTTCACCCACATGAAAAGCATCGGAATTGAG 62
336 TTCTCAATCTGTTTGTATGATCCGAGCATCTCAGGACACTAAAGATCATCTGATTTGGG 337
63 ATCGAGCTCAGAGACACCGGG 85

Db 336 ATCACACAACAGAGAACAGATGG 314

RESULT 6

US-08-486-663A-1/c
Sequence 1, Application US/0848663A
Patent No. 5968761
GENERAL INFORMATION:
APPLICANT: Rolfe, Mark
APPLICANT: Eckstein, Jens W.
APPLICANT: Draetta, Giulio
APPLICANT: Guillaume Cottarel
TITLE OF INVENTION: Ubiqutin Conjugating Enzymes
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,663A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-029CP2
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..441
US-08-486-663A-1

Query Match 17.6%; Score 30.2; DB 2; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

3 TTCTCAATACCGGAGAGGACAGAGCTATTTCACCCACATGAAAAGCATCGGAATTGAG 62
336 TTCTCAATCTGTTTGTATGATCCGAGCATCTCAGGACACTAAAGATCATCTGATTTGGG 337
63 ATCGAGCTCAGAGACACCGGG 85
336 ATCACACAACAGAGAACAGATGG 314

RESULT 7

US-08-903-396-1/c
Sequence 1, Application US/08903396
Patent No. 5968797
GENERAL INFORMATION:
APPLICANT: NI, ET AL.
TITLE OF INVENTION: Ubiqutin Conjugating Enzymes
FILING DATE: 7, 8 and 9

```

NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,396
FILING DATE: 22-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/464,342
FILING DATE: 5-JUN-1995
APPLICATION NUMBER: PCT/US95/01250
FILING DATE: 31-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-373
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-903-396-1

Query Match 17.6%; Score 30.2; DB 2; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACGACGCTATTTCACCCATGAAAGCATGGATTGAG 62
DB 396 TTCTCATCTGTTTGTGTGATCCGACATCTCAGGACCTAAAGATCATCTGATTGGG 337
QY 63 ATCGAGCTCAGAGGACACCCGG 85
DB 336 ATCAGACAGAGAGAACGATGG 314

RESULT 8
US-08-247-904B-1/c
Sequence 1, Application US/08247904B
Patent No. 5981699
GENERAL INFORMATION:
APPLICANT: Rolfe, Mark
APPLICANT: Eckstein, Jens W.
APPLICANT: Draetta, Giulio
TITLE OF INVENTION: Human Ubiquitin Conjugating Enzyme
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Eliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,904B
FILING DATE: 23-MAY-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-029.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..441
US-08-247-904B-1

Query Match 17.6%; Score 30.2; DB 2; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACGACGCTATTTCACCCATGAAAGCATGGATTGAG 62
DB 396 TTCTCATCTGTTTGTGTGATCCGACATCTCAGGACCTAAAGATCATCTGATTGGG 337
QY 63 ATCGAGCTCAGAGGACACCCGG 85
DB 336 ATCAGACAGAGAGAACGATGG 314

RESULT 9
US-08-895-601-4/c
Sequence 4, Application US/08895601
Patent No. 6060262
GENERAL INFORMATION:
APPLICANT: Beer-Romero, Peggy
APPLICANT: Strack, Peter J.
APPLICANT: Rolfe, Mark
TITLE OF INVENTION: REGULATION OF KAPPA B (1kB) DEGRADATION,
TITLE OF INVENTION: AND METHODS AND REAGENTS RELATED THERETO
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,601
FILING DATE: 16-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-096.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 4:
```

SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..441
US-08-695-601-4

Query Match 17.6%; Score 30.2; DB 3; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGGAATTGAG 62
DB 396 TTCTCTATCTGTTGGTAGATCCGAGCAATCTCAGGCACTAAAGATCATCTGGATTGGG 337
QY 63 ATCGAGCTCAGAGACACCGGG 85
DB 336 ATCACACACAGAGAACAGATGG 314

RESULT 10

US-08-767-942A-1/c
Sequence 1, Application US/08767942A
Patent No. 6068982

GENERAL INFORMATION:
APPLICANT: Rolfe, Mark
APPLICANT: Chiu, M. Isabel
APPLICANT: Berlin, Vivian
APPLICANT: Drazette, Veronique
APPLICANT: Guillaume, Cécile
TITLE OF INVENTION: UBIQUITIN CONUGATING ENZYMES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/767,942A
FILING DATE: 17-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV-029.04
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 444 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..441
US-08-767-942A-1

Query Match 17.6%; Score 30.2; DB 3; Length 444;
Best Local Similarity 60.2%; Pred. No. 0.25;

Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGGAATTGAG 62
DB 396 TTCTCTATCTGTTGGTAGATCCGAGCAATCTCAGGCACTAAAGATCATCTGGATTGGG 337
QY 63 ATCGAGCTCAGAGACACCGGG 85
DB 336 ATCACACACAGAGAACAGATGG 314

RESULT 11

US-08-425-299A-4/c
Sequence 4, Application US/08425299A
Patent No. 5726025

GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: King, Randall W.
APPLICANT: Peters, Jean-Michael
TITLE OF INVENTION: Assay and Reagents for Detecting Inhibitors
of Ubiquitin-Dependent Degradation of
TITLE OF INVENTION: Cell Cycle Regulatory Proteins
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,299A
FILING DATE: 20-APR-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMI-014
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 452 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..453
US-08-425-299A-4

Query Match 17.6%; Score 30.2; DB 1; Length 452;
Best Local Similarity 60.2%; Pred. No. 0.26;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGGAATTGAG 62
DB 405 TTCTCTATCTGTTGGTAGATCCGAGCAATCTCAGGCACTAAAGATCATCTGGATTGGG 336
QY 63 ATCGAGCTCAGAGACACCGGG 85
DB 345 ATCACACACAGAGAACAGATGG 323

RESULT 12

US-09-535-008-50/c

; Sequence 50, Application US/09535008
; Patent No. 6465629
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Tavligian, Sean V.
; APPLICANT: Teng, David H.-F.
; TITLE OF INVENTION: BRL1 IS A TUMOR SUPPRESSOR THAT IS MUTATED IN PROSTATE
; FILE REFERENCE: 2318-259
; CURRENT APPLICATION NUMBER: US/09/535,008
; CURRENT FILING DATE: 2000-03-23
; EARLIER APPLICATION NUMBER: U.S. 60/125,806
; EARLIER FILING DATE: 1998-03-23
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50
; LENGTH: 769
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-535-008-50

Query Match 17.0%; Score 29.2; DB 4; Length 769;
Best Local Similarity 62.2%; Pred. No. 0.72;
Matches 46; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 67 CAGCTCAGAGACACCGGCGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCT 126
DB 744 CAGCTAAAGTGACACAGGACCCCTTCCAGTGCGCCGACATGTGTGGAAGACT 685
QY 127 GGTGCTGAGACT 140
DB 684 CTCAGCTAGAGACT 671

RESULT 13
US-09-484-970B-119
; Sequence 119, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 119
; LENGTH: 2343
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186.1CB1
; LOCATION: 2323, 2332, 2337
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-119

Query Match 16.7%; Score 28.8; DB 4; Length 2343;
Best Local Similarity 52.5%; Pred. No. 1.6;
Matches 63; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 14 GGGAGAGGACAGAGCTATTCAGCCACATGAAAAGCATCGGAATTGATCGAGCTCA 73
DB 894 GAGGTAGGCCACAGCACCTCAGCAGAGAGATGAGCTGAGGGGACCACTCA 953
QY 74 GAGGACACCGGCGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCTTGAGCTGCC 133
DB 954 GACGGCTGTATATGCCCTGTTCTTGCGCTGGGTGGCGTGGGATCGCTAGCTTCC 1013

RESULT 14
US-08-839-008-8
; Sequence 8, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
; APPLICANT: Hurlie, Mark R
; APPLICANT: McDonnell, Peter C
; APPLICANT: McNulty, Dean E
; APPLICANT: Rosen, Craig A
; APPLICANT: Siemens, Ivo R
; APPLICANT: Young, Peter R
; APPLICANT: Yue, Jian-Li
; TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,008
; FILING DATE: 23-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/563,697
; FILING DATE: 28-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5096
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1480 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-839-008-8

Query Match 15.7%; Score 27; DB 2; Length 1480;
Best Local Similarity 62.7%; Pred. No. 5.6;
Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

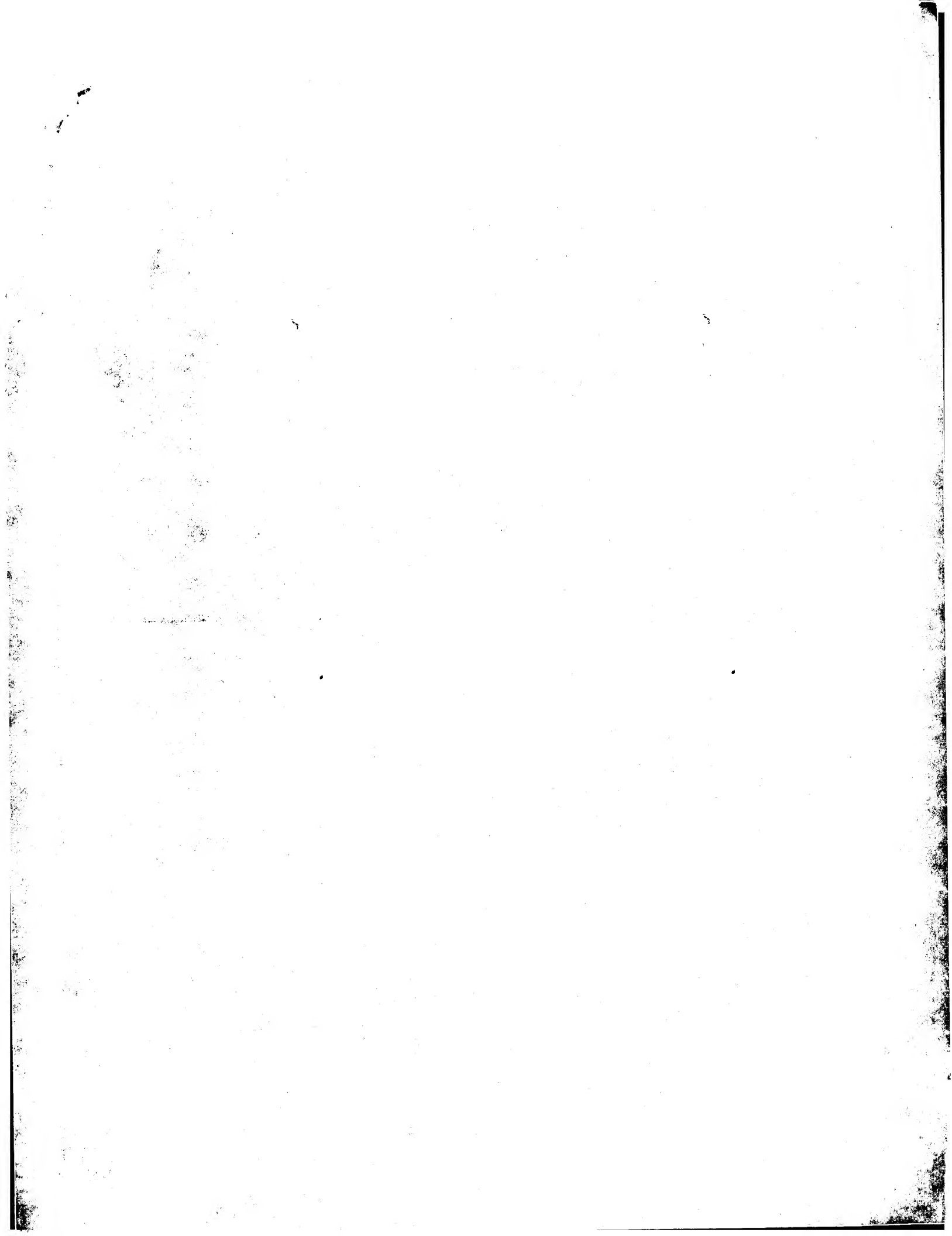
QY 73 AGAGACACCGGCGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCTTGCTGC 132
DB 1272 AGAAGAGACAGAGGCGCCCTTCCTTCAGAGAGCTTGTGTCTTCACCGGCCCA 1331
QY 133 CTGGGAC 139
DB 1332 CCAAGAC 1338

RESULT 15
US-08-839-008-1
; Sequence 1, Application US/08839008
; Patent No. 5916758
; GENERAL INFORMATION:
; APPLICANT: Hurlie, Mark R
; APPLICANT: McDonnell, Peter C
; APPLICANT: McNulty, Dean E
; APPLICANT: Rosen, Craig A
; APPLICANT: Siemens, Ivo R
; APPLICANT: Young, Peter R

APPLICANT: Yue, Tian-Li
 TITLE OF INVENTION: Smooth Muscle Cell-Derived Migration Factor
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SmithKline Beecham Corporation
 STREET: 709 Swedeland Road
 CITY: King of Prussia
 STATE: PA
 COUNTRY: USA
 ZIP: 19406
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/839,008
 FILING DATE: 23-APR-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/563,697
 FILING DATE: 28-NOV-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Baumeister, Kirk
 REGISTRATION NUMBER: 33,833
 REFERENCE/DOCKET NUMBER: P50384
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-5096
 TELEFAX: 610-270-5090
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1537 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 US-08-839-008-1

Query Match 15.7%; Score 27; DB 2; Length 1537;
 Best Local Similarity 62.7%; Pred. No. 5.7;
 Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 73 AGAGACACCGGCGCCCTTCCACCTTCCAGAGAGCTTTGTAATTTGCAATCTGGCTGC 132
 DB 1305 AGAAGAGAAGAGAGCGCCCTTCTTCTTCCAGAGAGCTTTGTAATTTGCAATCTGGCTGC 132
 QY 133 CTGGGAC 139
 DB 1365 CCAGGAC 1371

Search completed: April 15, 2003, 22:46:48
 Job time : 13.1771 secs



US-09-960-352-5216
; Sequence 5216, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5216
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 23-BOWMS1-014-Q1-E1-F3
US-09-960-352-5216

Query Match 24.3%; Score 41.8; DB 10; Length 466;
Best Local Similarity 73.6%; Pred. No. 2.8e-06;
Matches 67; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

QY 81 CCGGGCGCCCTTCACCTTCCAGAGAGCTTGTATCTTCGATCTGGCTGCGGACT 140
DB 1 CCGGGCGCCCTTCACCTTCCAGAGAGCTTGTATCTTCGATCTGGCTGCGGACT 58

QY 141 TCCCTTAGGAGTAAACAATACATTAAGCA 171
DB 59 TTCCTCAGGATGAACAATGTGCATATAGCA 89

RESULT 3
US-10-043-487-198/c
; Sequence 198, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEGRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043.487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 198
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Shigella flexneri
US-10-043-487-198

Query Match 17.6%; Score 30.2; DB 9; Length 433;
Best Local Similarity 60.2%; Pred. No. 0.094;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAGAAGCATCGAATTGAG 62
DB 385 TTCTCTATCTGTTTGTAGATCCGAGCATCTCAGGCACTAAAGATATCTGATTGGG 326

QY 63 ATGCAAGCTCAGAGAGACCCGGG 85
DB 325 ATCACACAACAGAGAACGATGG 303

RESULT 4
US-09-833-381-896/c
; Sequence 896, Application US/09833381
; Patent No. US20020132090A1

; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833.381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 896
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1946)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-896

Query Match 17.6%; Score 30.2; DB 10; Length 1946;
Best Local Similarity 60.2%; Pred. No. 0.18;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAGAAGCATCGAATTGAG 62
DB 840 TTCTCTATCTGTTTGTAGATCCGAGCATCTCAGGCACTAAAGATATCTGATTGGG 781

QY 63 ATGCAAGCTCAGAGAGACCCGGG 85
DB 780 ATCACACAACAGAGAACGATGG 758

RESULT 5
US-10-026-188-3/c
; Sequence 3, Application US/10026188
; Patent No. US20020164645A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Zhang, Yifeng
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
; FILE REFERENCE: 02307E-114910US
; CURRENT APPLICATION NUMBER: US/10/026.188
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/259,379
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 249487
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse genomic region containing 1trpc5
US-10-026-188-3

Query Match 17.4%; Score 30; DB 9; Length 249487;
Best Local Similarity 53.4%; Pred. No. 1.7;
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 35 CAGCCATATAAAGACATCGAAATTGATCGAGCTCAGAGACACCGGCGCCCTTC 94
DB 127121 CAGCCATAGCATCTTGATTGTCTGAGAGTGCAGCCGCGACAGCACCGTCTCTGCC 127062

QY 95 CACCTTCAAGAGAGCTTTGATCTTCGATCTGCGCTGCGGATCTCCCTTAGCGAG 152
DB 127061 CCAAAACAGTGTAGTACATAGCAGCTTCAGGCGCGCGGTGATCTCCCAAGTCAG 127004

RESULT 6
US-09-770-445-525/c

Sequence 525, Application US/09770445
Patent No. US2002023281A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matheuw, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Moesener, Jeffrey P.
APPLICANT: Heas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kriker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 202305 (PARA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 525
LENGTH: 881
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-445-525

Query Match 17.2%; Score 29.6; DB 10; Length 881;
Best Local Similarity 50.7%; Pred. No. 0.22;
Matches 71; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 6 TCATATCCGGAGAGGACGACGACTTATTCAGCCACATGAAAGCATGGAAATGAGATC 65
DB 270 TCAGGCGAAGAGGAGACGAGATGATTCATCTGCTCGACCTTACGATGCGCTTC 211
QY 66 GCAGCTGAGAGGACACCGGGGCGCCCTTCACCTTCCAGAGAGCTTGATTTGATC 125
DB 210 TCAGCGTAGATCCGAGACCATCATATTAACCTTTCGATGAGCTTCTTCACTCAAT 151
QY 126 TGGCTGCTGGGACTTCCT 145
DB 150 TTCTGCTTCTTCTTCTT 131

RESULT 7
US-09-917-800A-1389/C

Sequence 1389, Application US/09917800A
Patent No. US20020119462A1

GENERAL INFORMATION:

APPLICANT: Mendrick, Donna

APPLICANT: Porter, Mark

APPLICANT: Johnson, Kory

APPLICANT: Castle, Arthur

APPLICANT: Blaschhoff, Michael

APPLICANT: Gene Logic, Inc.

TITLE OF INVENTION: Molecular Toxicology Modeling

FILE REFERENCE: 44921-5038-US

CURRENT APPLICATION NUMBER: US/09/917,800A

CURRENT FILING DATE: 2001-07-31

PRIOR APPLICATION NUMBER: US 60/222,040

PRIOR FILING DATE: 2000-07-31

PRIOR APPLICATION NUMBER: US 60/222,880

PRIOR FILING DATE: 2000-11-02

PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1389
LENGTH: 952
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 L19698
US-09-917-800A-1389

Query Match 16.4%; Score 28.2; DB 10; Length 952;
Best Local Similarity 68.4%; Pred. No. 0.8;
Matches 39; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 109 CTTTGATTTCTTGATCTGGCTGCTGGGACTTCCTTGAAGCATTAATACAT 165
DB 827 CTTTGCTTCTTCATCTTCTGCTGCTGATATTCCTCTTAATCAAAAATAACCT 771

RESULT 8

US-09-877-065-2

Sequence 2, Application US/09877065

Patent No. US2002005190A1

GENERAL INFORMATION:

APPLICANT: OPLE, ERIC

APPLICANT: MCLACHLAN, KAREN

APPLICANT: HEARD, CHERYL J.

TITLE OF INVENTION: NOVEL GENE TARGETS AND LIGANDS THAT BIND THERETO FOR

TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF OVARIAN CARCINOMAS

FILE REFERENCE: 037003-0280631

CURRENT APPLICATION NUMBER: US/09/877,065

PRIOR APPLICATION NUMBER: 60/210,451

PRIOR FILING DATE: 2000-06-09

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 1257

TYPE: DNA

ORGANISM: Homo sapiens

US-09-877-065-2

Query Match 16.3%; Score 28; DB 10; Length 1257;
Best Local Similarity 50.8%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 36 AGCCACATGAAGACATCGAATTCAGATCGAGCTCAGAGACACCGGCGCCCTTC 95
DB 1039 AGCCAGAGAGGAGACTGATTCCTCAGTCCACAGCTCAGACAGACACCGGCGCC 1098
QY 96 ACCCTCAAGAGCTTGATTTCTTGATCTGGCTGCTGCTGCGACTTCCTTGAAGCA 155
DB 1099 ACCATCAAGCCCTTTATTTATTTATTAAGTGACTTTTATTTATTTATTTT 1158
QY 156 ACAATACATTA 167
DB 1159 AAAAAAGAAAA 1170

RESULT 9
US-09-878-574-12958/C
Sequence 12958, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878,574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/333,535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 12958
LENGTH: 180
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: 701066451H1
US-09-878-574-12958

Query Match 16.2%; Score 27.8; DB 10; Length 180;
Best Local Similarity 57.5%; Pred. No. 0.56; Mismatches 37; Indels 0; Gaps 0;
Matches 50; Conservative 0;

QY 55 GAATTGAGATGCACTCAGAGGACCGGGCCCTTCACCTTCAGAGAGCTTGT 114
DB 171 GCAAGAGATGACACCTGATCTCAGAGAGGCGCTTGATCTTCTCCACCTTCACT 112

QY 115 ATTCTGATCTGCTGCTGCTGAGCTT 141
DB 111 TTGCTGATCTGCTGCTGCTGATGCACTT 85

RESULT 10
US-09-919-497-39
Sequence 39, Application US/09919497
Patent No. US2002010662A1
GENERAL INFORMATION:
APPLICANT: Mutter, George L.
TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
FILE REFERENCE: B0801/77225
CURRENT APPLICATION NUMBER: US/09/919,497
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/221,735
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn version 3.0
SEQ ID NO 39
LENGTH: 1480
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-497-39

Query Match 15.7%; Score 27; DB 10; Length 1480;
Best Local Similarity 62.7%; Pred. No. 2.8; Mismatches 25; Indels 0; Gaps 0;
Matches 42; Conservative 0;

QY 73 AGAGGACACCGGGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCTGAGCTGCTGC 132
DB 1272 AGAAGAGAACAGAGGCGCCCTTCCTTCCTCCAGAGAGCTTGTGTTCTTCACCGGCCCA 1331

QY 133 CTGGGAC 139
DB 1332 CCAGGAC 1338

RESULT 11
US-09-880-107-2196
Sequence 2196, Application US/09880107
Patent No. US20020142981A1

GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Schief, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2196
LENGTH: 1480
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 L33799
US-09-880-107-2196

Query Match 15.7%; Score 27; DB 10; Length 1480;
Best Local Similarity 62.7%; Pred. No. 2.8; Mismatches 25; Indels 0; Gaps 0;
Matches 42; Conservative 0;

QY 73 AGAGGACACCGGGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCTGAGCTGCTGC 132
DB 1272 AGAAGAGAACAGAGGCGCCCTTCCTTCCTCCAGAGAGCTTGTGTTCTTCACCGGCCCA 1331

QY 133 CTGGGAC 139
DB 1332 CCAGGAC 1338

RESULT 12
US-09-822-849A-273
Sequence 273, Application US/09822849A
Patent No. US20020045170A1
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakara
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6403
CURRENT APPLICATION NUMBER: US/09/822,849A
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/195,582
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 598
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 273
LENGTH: 1542
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-849A-273

Query Match 15.7%; Score 27; DB 10; Length 1542;
Best Local Similarity 62.7%; Pred. No. 2.9; Mismatches 25; Indels 0; Gaps 0;
Matches 42; Conservative 0;

QY 73 AGAGGACACCGGGCCCTTCACCTTCAGAGAGCTTGTATCTTGATCTGAGCTGCTGC 132
DB 1327 AGAAGAGAACAGAGGCGCCCTTCCTTCCTCCAGAGAGCTTGTGTTCTTCACCGGCCCA 1386

QY 133 CTGGGAC 139

Db 1387 CCAGGAC 1393

RESULT 13

US-09-925-301-440
Sequence 440, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925.301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 440
LENGTH: 1580
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (873)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-440

Query Match 15.7%; Score 27; DB 10; Length 1580;
Best Local Similarity 62.7%; Pred. No. 2.9;
Matches 42; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 73 AGAGACACGGGGCCCTTCACCTTCAGAGAGCTTTATTTGCACTGGCGTCC 132

Db 1341 AGAAGAAACGAGAGCCCGCTTCCTCCAGAGAGCTTTGCTTCACCGCCCA 1400

Qy 133 CTGGGAC 139

Db - 1401 CCAGGAC 1407

RESULT 14

US-10-028-072-79
Sequence 79, Application US/10028072
Publication No. US20030004311A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerilsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028.072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26

PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334

;; PRIOR FILING DATE: 1997-12-11
;; PRIOR APPLICATION NUMBER: 60/069694
;; PRIOR FILING DATE: 1997-12-15
;; PRIOR APPLICATION NUMBER: 60/073320
;; PRIOR FILING DATE: 1998-01-23
;; PRIOR APPLICATION NUMBER: 60/073612
;; PRIOR FILING DATE: 1998-02-04
;; PRIOR APPLICATION NUMBER: 60/074086
;; PRIOR FILING DATE: 1998-02-09
;; PRIOR APPLICATION NUMBER: 60/074092
;; PRIOR FILING DATE: 1998-02-09
;; PRIOR APPLICATION NUMBER: 60/077791
;; PRIOR FILING DATE: 1998-03-12
;; PRIOR APPLICATION NUMBER: 60/078910
;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/079294
;; PRIOR FILING DATE: 1998-03-25
;; PRIOR APPLICATION NUMBER: 60/079663
;; PRIOR FILING DATE: 1998-02-27
;; PRIOR APPLICATION NUMBER: 60/079728
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/080165
;; PRIOR FILING DATE: 1998-03-31
;; PRIOR APPLICATION NUMBER: 60/081203
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081229
;; PRIOR FILING DATE: 1998-04-09
;; PRIOR APPLICATION NUMBER: 60/081695
;; PRIOR FILING DATE: 1998-04-14
;; PRIOR APPLICATION NUMBER: 60/081817
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/081818
;; PRIOR FILING DATE: 1998-04-15
;; PRIOR APPLICATION NUMBER: 60/082999
;; PRIOR FILING DATE: 1998-04-24
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/083545
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085149
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086414
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086430
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088730
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088741
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10

;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089539
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090538
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07

Query Match 15.7%; Score 27; DB 9; Length 2714;
Best Local Similarity 56.0%; Pred. No. 3.7;
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 20 GGCACAGCTATTTCAGCCACATGAAAGCATGGCAATTGAGATCCAGCTCAGAGAC 79
DB 2204 GGAAGATGGCTGTGTTCAGCCATTACCTCTGAGATGGGAGATCTCTCCAGAAC 2263
DB 2264 ACATGAGAGCTGCTCTTCGATCCAGCAGGT 2294

RESULT 15
US-10-121-049-79
; Sequence 79, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Guiney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumaas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 79
; LENGTH: 2714
; TYPE: DNA
; ORGANISM: Homo Sapien

US-10-121-049-79

Query Match 15.7%; Score 27; DB 9; Length 2714;
 Best Local Similarity 56.0%; Pred. No. 3.7;
 Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 20 GGCACAGACTATTTCAGCCACATGAAAAGCATCGGAATTGAGATGCAGCTCAGAGGAC 79
 Db 2204 GGAAGATGGCTTGTTCAGCCCATTCACCTCTGAGGATGTGGCAGTCTCTTCAAGAAC 2263
 Qy 80 ACCGGGCGCCCTTCCACCTTCCAAAGAGCT 110
 Db 2264 ACATGAGCTGCTTCTGATCCCAAGCAGGT 2294

Search completed: April 16, 2003, 01:14:11
 Job time : 80.1865 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:26:41 ; Search time 45.8981 Seconds
(without alignments)
8439.207 Million cell updates/sec

Title: US-09-647-019-6

Perfect score: 172
Sequence: 1 ggtctcaataacgggagag.....taacaatacataaagcag 172

Scoring table: IDENTITY NUC
Gapop 10-0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID82/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	172	100.0	887	20	AAK90904
2	171	99.4	428	21	AAAC01483
3	171	99.4	886	24	AAZ727216
4	95.2	55.3	778	20	AAK90903
5	31.2	18.1	86574	24	ABK83560
6	30.8	17.9	1563	21	AA26372
7	30.2	17.6	444	16	AAQ97831
8	30.2	17.6	444	17	AAK39685
9	30.2	17.6	444	18	AAK79814

C 10	30.2	17.6	444	20	AA225304	Human ubiquitin co
C 11	30.2	17.6	444	20	AA227558	Human ubiquitin co
C 12	30.2	17.6	444	20	AAK78496	Human UBC4 cDNA.
C 13	30.2	17.6	444	20	AAV82865	Ubiquitin conjugat
C 14	30.2	17.6	444	21	AA61613	cDNA encoding huma
C 15	30.2	17.6	452	17	AA741769	Human ubiquitin-co
C 16	30.2	17.6	509	24	ABK84311	Human cDNA differe
C 17	30.2	17.6	1732	22	AAK14153	Human cDNA sequenc
C 18	30.2	17.6	2410	24	AB199458	Human cDNA sequenc
C 19	30.2	17.6	2526	22	AAH14290	Mouse ischaemic co
C 20	30.2	17.6	5069	22	AAH14699	Human cDNA sequenc
C 21	30.2	17.6	5076	22	ABAI4700	Human nervous syst
C 22	30	17.4	249487	24	ABN85733	Human nervous syst
C 23	29.8	17.3	1098	23	AA570300	Mouse genomic regi
C 24	29.8	17.3	1529	21	AAK59838	DNA encoding novel
C 25	29.8	17.3	6288	22	AAK45090	Human secreted pro
C 26	29.6	17.2	881	24	ABN98757	cDNA encoding nove
C 27	29.6	17.2	948	21	AAK50935	Arabidopsis thaila
C 28	29.6	17.2	950	21	AAK44655	Arabidopsis thaila
C 29	29.4	17.1	2672	22	AAI07039	Human reproductive
C 30	29.4	17.1	2672	22	AAI62717	Human breast or ov
C 31	29.2	17.0	337	24	ABN2475	Human ORFX polynuc
C 32	29.2	17.0	769	21	AAK58894	Human tumor suppr
C 33	29.2	17.0	795	22	ABAI8324	Human nervous syst
C 34	29.2	17.0	1522	24	ABL57467	Human protein phos
C 35	29.2	17.0	2631	23	AB103923	Drosophila melanog
C 36	29.2	17.0	7889	23	AB103922	Drosophila melanog
C 37	29.2	17.0	27289	22	AAK34624	Human DNA for a no
C 38	28.8	16.7	743	21	AAK96915	Human secreted pro
C 39	28.8	16.7	2255	21	AAK52664	Eosinophil activat
C 40	28.8	16.7	2309	21	AAK27137	Human inflammation
C 41	28.8	16.7	4431	22	AAH27100	Human GTPase activ
C 42	28.6	16.6	518	20	AAK23399	Human Delta7-steri
C 43	28.6	16.6	1314	23	AAK73940	DNA encoding novel
C 44	28.6	16.6	2818	23	AB123165	Drosophila melanog
C 45	28.6	16.6	6101	23	AB123164	Drosophila melanog

ALIGNMENTS

RESULT 1	AAK90904	standard; cDNA; 887 bp.
ID	AAK90904;	
AC	AAK90904;	
XX	17-JAN-2000 (first entry)	
DT	XX	
DE	XX	
CDNA	encoding human chisel (Csl) gene.	
XX	XX	
KW	Chisel gene; Csl; EF-Hand protein super family; muscle development;	
KW	heart/skeletal muscle cell development; signalling pathway; regulation;	
KW	Xq21.3-q22; adaptive process; muscle homeostasis; skeletal myopathy;	
KW	detection; diagnosis; prophylaxis; treatment; cardiac hypertrophy;	
KW	muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;	
KW	Becker's myotonic dystrophy; heart failure; differentiation; exon;	
XX	gene therapy; transgenic animal; drug screening; 89.	
OS	Homo sapiens.	
XX	XX	
FH	Key	Location/Qualifiers
FT	exon	1..172
FT	FT	/*tag= a
FT	FT	/label= Exon_1
FT	FT	/note= "Corresponds to residues 19497-19327 of
FT	FT	human cosmid clone U22804"
FT	FT	173..229
FT	FT	/*tag= b
FT	FT	/label= Exon_2
FT	FT	/note= "Corresponds to residues 15687-15631 of
FT	FT	human cosmid clone U22804"
FT	FT	185..451
FT	FT	CDS

```

FT FT      /*tag= c
FT FT      /product= "Chisel (Cal) protein"
FT FT      /note= "Expressed predominantly in heart muscles"
FT FT      exon
FT FT      230..316
FT FT      /tag= d
FT FT      /label= Exon_3
FT FT      /note= "Corresponds to residues 5220-5134 of
FT FT      human cosmid clone U228D4"
FT FT      exon
FT FT      317..465
FT FT      /*tag= e
FT FT      /label= Exon_4
FT FT      /note= "Corresponds to residues 35384-35236 of
FT FT      human cosmid clone U112B8"
FT FT      exon
FT FT      466..887
FT FT      /tag= f
FT FT      /label= Exon_5
FT FT      /note= "Corresponds to residues 4101-3680 of
FT FT      human cosmid clone U112B8"
XX XX      WO9950410-A1.
XX XX      07-OCT-1999.
XX XX      26-MAR-1999; 99WO-AU00220.
XX XX      27-MAR-1998; 98AU-0002634.
XX XX      (CHAN-) CHANG CARDIAC RES INST VICTOR.
XX XX      (GEO) GEN HOSPITAL CORP.
XX XX      (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX XX      Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
XX XX      WPI; 1999-610852/52.
XX XX      P-PSDB; AAY28651.
XX XX      Isolated nucleic acids encoding chisel, used to develop products for
XX XX      treating cardiomyopathy, cardiac hypertrophy, heart failure and
XX XX      muscular myopathies -
XX XX      Claim 8; Page 149-150; 157pp; English.
XX XX      The present sequence is the cDNA encoding the human chisel gene (Cal)
XX XX      that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.
XX XX      It is a member of the EF-Hand protein super family and is involved in
XX XX      signalling pathways. It is predominantly expressed in heart and skeletal
XX XX      muscles and is activated after the differentiation of cells. Csl
XX XX      functions in regulation aspects of differentiation or adaptive processes
XX XX      that maintain muscle homeostasis. This sequence can be used in the
XX XX      detection, diagnosis, prophylactic and therapeutic treatment and functional
XX XX      such as those involving aberrant muscle cell development and functional
XX XX      activity. It is also used in the treatment of muscular and myotonic
XX XX      dystrophies, skeletal myopathies such as Duchenne muscular dystrophy and
XX XX      Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,
XX XX      myocarditis, myofiber atrophy, etc. The Csl gene sequence can also be
XX XX      used in gene therapy, for the production of transgenic animals and for
XX XX      drug screening.
SQ SQ      Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;
XX XX      Query Match 100.0%; Score 172; DB 20; Length 887;
XX XX      Best Local Similarity 100.0%; Pred. No. 1.8e-51;
XX XX      Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY QY      1 GGTTTCATATCCGGAGAGGCAACAGGCTATTTCCAGCATGAAACATCGAATTG 60
QY      1 GGTTCATATCCGGAGAGGCAACAGGCTATTTCCAGCATGAAACATCGAATTG 60
DB DB      61 AGATCGCAGCTCAGAGAACACCGGGGCGCCCTTCCACTTCCAGAGCGTTTGATCTT 120
DB      61 AGATCGCAGCTCAGAGAACACCGGGGCGCCCTTCCACTTCCAGAGCGTTTGATCTT 120
QY QY      121 GCATCTGGCTGCTGGGACTTCCCTTAGGCGATGAACAAATACATATAAGACG 172

```

Db 121 GCATCTGGCTGCTGGAGCTTCCCTTAGGACAGTAACAATAACATTAAGCAG 172

RESULT 2
AAC01483
ID AAC01483 standard; cDNA; 428 BP.
XX AAC01483;
XX
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST. SEQ ID NO: 1481.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
XX Dunas Milne Edwards J, Nucleart A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX P-PDB; AAG01477.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 1481; 71pp + CD-ROM; English.
XX
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX cDNAs encoding secreted proteins. An ORF has been identified within the
XX sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
XX derived from 30 different tissues. EST sequences usually correspond
XX mainly to the 3' untranslated region (UTR) of the mRNA because they are
XX often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
XX well suited for isolating cDNA sequences derived from the 5' ends of
XX mRNAs and even in those cases where longer cDNA sequences have been
XX obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
XX cDNAs with intact 5' ends and can therefore be used to obtain full length
XX cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
XX gene therapy and chromosome mapping procedures. They are used to obtain
XX upstream regulatory sequences and to design expression and secretion
XX vectors.
XX
XX Sequence 428 BP; 123 A; 111 C; 104 G; 90 T; 0 other;
XX
XX
XX Query Match 99.4%; Score 171; DB 21; Length 428;
XX Best Local Similarity 100.0%; Pred. No. 3e-51;
XX Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX 2 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCACAATGAAGAATCGGAATTGA 61
XX 65 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCACAATGAAGAATCGGAATTGA 124
XX
XX 62 GATCGAGTTCAGAGACACCGGGCCCTCCACCTTCCAAAGAGCTTGATTTCTTG 121
XX 125 GATCGAGTTCAGAGACACCGGGCCCTTCCACCTTCCAAAGAGCTTGATTTCTTG 184
XX
XX 122 CATCTGGCTGCTGGAGCTTCCCTTAGGACAGTAACAATAACATTAAGCAG 172
XX
XX 185 CATCTGGCTGCTGGAGCTTCCCTTAGGACAGTAACAATAACATTAAGCAG 235
XX


```

RESULT 3
AAD27216
ID AAD27216 standard; DNA; 886 BP.
XX
AC AAD27216;
XX
DT 09-APR-2002 (first entry)
XX
DE Human 66214 EST clone DNA.
XX
KM Human; congestive heart failure; dilative cardiomyopathy; sudden death;
KM hypertrophic cardiomyopathy; ischemic cardiomyopathy; rhythm disorder;
KM heart muscle disease; conduction disorder; coronary heart disease;
KM systemic arterial hypertension; pulmonary hypertension; endocarditis;
KM pulmonary heart disease; valvular heart disease; pericardial disease;
KM congenital heart disease; gene therapy; syncope; transgenic animal;
KM expressed sequence tag; EST; clone 66214; db.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 184..450
FT /tag= a
FT /product= "Human 66214 protein"
FT misc_feature 298..588
FT /tag= b
FT /note= "66214 cDNA fragment"
FT polyA_signal 857..862
FT /tag= c
XX
PN MO200192567-A2.
XX
PD 06-DEC-2001.
XX
PE 30-MAY-2001; 2001MO-EP06165.
XX
PR 30-MAY-2000; 2000US-207400P.
XX
PA (MED1-) MEDIGENE AG.
XX
PI Bunk D, Reuner B, Beck J, Henkel T;
XX
DR WPI; 2002-122073/16.
XX
DR P-PSDB; AAE16632.
XX
PT Identifying a subject at risk for a heart disease e.g. congestive heart
PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
PT the polypeptide expressed by genes abnormally expressed in heart tissue
XX
XX
PS Claim 2a; Fig 9b; 154bp; English.
XX
CC The patent discloses novel target genes abnormally expressed in heart
CC tissues and their corresponding proteins. The invention also relates to
CC methods for assessing the expression level of these genes. The method
CC is used for testing the predisposition of mammals and preferably humans
CC for a heart disease or for an acute state of such a disease. It is also
CC useful to treat diseases of the heart such as congestive heart failure,
CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischemic cardio-
CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
CC syncope and sudden death, coronary heart disease, systemic arterial
CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
CC heart disease, congenital heart disease, pericardial disease and
CC endocarditis. Sequences of the invention are also used in gene therapy.
CC A transgenic non-human mammal comprising the sequences of the invention
CC are useful for the development for medications for the treatments of
CC heart diseases. The present DNA sequence is expressed sequence tag
CC (EST) 66214 clone.
XX
SQ Sequence 886 BP; 278 A; 172 C; 191 G; 245 T; 0 other;
Query Match 99.4%; Score 171; DB 24; Length 886;

```

```

Best Local Similarity 100.0%; Pred. No. 46-51;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GTTCTCAATACCGGAGACAGACACTTATTGACCATGTAAGCAATCGAATTGA 61
DB 1 GTTCTCAATACCGGAGACAGACACTTATTGACCATGTAAGCAATCGAATTGA 60
QY 62 GATCGACGCTCAGAGACACCGGAGCGCCCTTCCACCTCCAGAGACTTGTATCTTG 121
DB 61 GATCGACGCTCAGAGACACCGGAGCGCCCTTCCACCTCCAGAGACTTGTATCTTG 120
QY 122 CATCTGCGCTGCTGGGACTTCCCTTAGGCAGTAAACAAATACATTAAGCAG 172
DB 121 CATCTGCGCTGCTGGGACTTCCCTTAGGCAGTAAACAAATACATTAAGCAG 171
RESULT 4
AAX90903
ID AAX90903 standard; cDNA; 778 BP.
XX
AC AAX90903;
XX
DT 17-JAN-2000 (first entry)
XX
DE cDNA encoding murine chisel (Csl) gene.
XX
KM Chisel gene; Csl; EF-Hand protein super family; muscle development;
KM heart/skeletal muscle cell development; signalling pathway; murine;
KM X chromosome; regulation; adaptive process; muscle homeostasis;
KM detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
KM Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
KM differentiation; gene therapy; transgenic animal; drug screening; ss.
XX
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 199..456
FT /tag= a
FT /product= "Chisel (Csl) protein"
FT /note= "Expressed especially in heart muscles"
XX
PN MO9950410-A1.
XX
PD 07-OCT-1999.
XX
PE 26-MAR-1999; 99WO-AU00220.
XX
PR 27-MAR-1998; 98AU-0002634.
XX
PA (CHAN-) CHANG CARDIAC RES INST VICTOR.
PA (GEHO) GEN HOSPITAL CORP.
PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
XX
DR WPI; 1999-610852/52.
XX
DR P-PSDB; AAT28650.
XX
PT Isolated nucleic acids encoding chisel, used to develop products for
PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
PT muscular myopathies -
XX
PS Claim 3; Page 148; 157pp; English.
XX
CC The present sequence is the cDNA encoding the murine chisel gene (Csl)
CC that is mapped to the mouse X chromosome. It is a member of the EF-Hand
CC protein super family and is involved in signalling pathways. It is
CC predominantly expressed in heart and skeletal muscles and is activated
CC after the differentiation or adaptive processes that maintain muscle
CC homeostasis. This sequence can be used in the detection, diagnosis,
CC prophylactic and therapeutic treatment of diseases such as those

```

CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Cal gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.

XX Sequence 778 BP; 231 A; 166 C; 179 G; 202 T; 0 other;

SQ Query Match 55.3%; Score 95.2; DB 20; Length 778;
 Best Local Similarity 76.5%; Pred. No. 6,8e-24;
 Matches 130; Conservative 0; Mismatches 36; Indels 2; Gaps 1;

QY 5 CTCATATCCGGAGAGGACAGAGCTATTTCAGCCATGTAAGAAAGCATCGAATTGAGAT 64
 DB 11 CTGAGAGAGAGACAGAGACTCCAGCTATTTCAGCCATGTAAGAAAGCATCGAATTGAGAT 70
 QY 65 CGCAGCTCAGAGAGACACCGGCGCCCTTCCACCTTCCAGAGAGC--TTTGTATTCTTGC 122
 DB 71 CCCGCTCAGAGAGACACCGGAGTTCCTTCTATCTGTAAGCGCTTTTGTGTTTTC 130
 QY 123 ATCTGGCTGCTGGGAGCTTCCCTTAGGAGATTAACAATATCATTAAGCAG 172
 DB 131 ACCTGGCGGCTGGGAGCTGTCTCTCAGGAGTAACCAATCCAGAGAGCAG 180

RESULT 5
 ABRK3560/c
 ID ABRK3560 standard; cDNA; 86574 BP.

AC ABRK3560;

XX 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #131.

DE Human; ss; granulocytic cell; DNA chip; bacterial infection;
 XX viral infection; parasitic infection; protozoal infection;
 KW fungal infection; sterile inflammatory disease; psoriasis;
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KW cardiac reperfusion injury; renal reperfusion injury; AIDS;
 KW adult respiratory distress syndrome; inflammatory bowel disease;
 KW Crohn's disease; ulcerative colitis; periodontal disease;
 KW granulocyte activation; chronic inflammation; allergy.

XX Homo sapiens.

XX WO200228999-A2.

XX 11-APR-2002.

XX 03-OCT-2001; 2001WO-US30821.

XX 03-OCT-2000; 2000US-237189P.

XX (GENE-) GENE LOGIC INC.

XX Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;

XX WPI; 2002-435328/46.

PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -

PS Claim 1; SEQ ID No 131; 114pp; English.

CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.

CC Also included are modulating (M2) GA by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 86574 BP; 22071 A; 20398 C; 21552 G; 22553 T; 0 other;

Query Match 18.1%; Score 31.2; DB 24; Length 86574;
 Best Local Similarity 63.2%; Pred. No. 5;
 Matches 48; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 81 CCGGGCGCCCTTCCACCTTCCAGAGAGCTTGTATTCTTGATCTGGCTGGGACT 140
 DB 51402 CAGAGAGGGCTTCCAGCTTCTTAGATCTTGTCTTCTTCTAGTACACAGAGAACT 51343
 QY 141 TCCTTAGGAGTA 156
 DB 51342 TGACAGAGTGA 51327

RESULT 6

XX AAA26372 standard; cDNA; 1563 BP.

XX AAA26372;

XX 29-JUN-2000 (first entry)

XX Human secreted protein gene 27 SEQ ID NO:37.

DE Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW antiHIV; antiinflammatory; nootropic; neuroprotective; antiallergic;
 KW osteopathic; antiarthritic; antibacterial; antidiabetic; antisthma;
 KW antipsoriatic; cardiac; gene therapy; cancer; neurological disorder;
 KW immune disease; inflammation; blood disorder; tumour; ss.

XX Homo sapiens.

XX WO200006698-A1.

XX 10-FEB-2000.

XX 29-JUL-1999; 99WO-US17130.

XX 30-JUL-1998; 98US-0094657.

XX 05-AUG-1998; 98US-0095486.

KW immunosuppression related disorder; hymenoptera venom hypersensitivity;
 KW aspirin-induced asthma; Alzheimer's disease; atrophying skeletal muscle;
 KW African swine fever virus; apoptotic cell death; human; ss.
 XX
 OS Homo sapiens.
 XX
 PN W09623410-A1.
 XX
 PD 08-AUG-1996.
 XX
 PF 31-JAN-1995; 95WO-US01250.
 XX
 PR 31-JAN-1995; 95WO-US01250.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Adams MD, Gentz R, Ni J;
 XX
 DR WPI; 1996-371042/37.
 DR P-PSDB; AAW05313.
 XX
 PT Ubiquitin Conjugating Enzyme 7, 8 and 9 polypeptide(s) - useful for
 PT the treatment of e.g. the proliferation of malignant cells,
 PT Alzheimer's disease and immunological disorders
 XX
 PS Claim 1; Page 52; 85pp; English.
 XX
 CC AAT93695-T33697 represent the coding sequences for the ubiquitin
 CC conjugating enzymes (UCE) 7, 8, and 9 respectively. Mammalian cells
 CC contain two distinct proteolytic pathways, one of which is ubiquitin
 CC dependent. This proteolytic pathway requires the post-translational
 CC attachment of ubiquitin to other proteins using UCEs, such as the
 CC encoded proteins. Ubiquitin-protein conjugation is highly selective and
 CC is required for a large variety of cellular functions, including DNA
 CC repair, cell viability, heat shock resistance, and cell cycle
 CC progression. The UCEs 7, 8, and 9 may also play a role in selective
 CC protein degradation in human cells. The UCE proteins encoded by these
 CC sequences can be used to treat patients that require one of the UCEs, or
 CC need one of the UCEs inhibited. The three UCEs can also be used for the
 CC treatment of the proliferation of malignant cells. The encoded proteins
 CC can also be used to treat immunosuppression related disorders, such as
 CC AIDS, or to regulate immunological disorders such as hymenoptera venom
 CC hypersensitivity or aspirin-induced asthma. Antagonists against the
 CC UCEs can be used to treat disorders such as Alzheimer's disease,
 CC atrophying skeletal muscle, African swine fever virus, and apoptotic cell
 CC death.
 XX
 SQ Sequence 444 BP; 135 A; 95 C; 91 G; 123 T; 0 other;
 XX
 Query Match 17.6%; Score 30.2; DB 17; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 XX
 QY 3 TTCTCATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAGAATCGAATTGAG 62
 Db 396 TTCTCTATCTGTTTGTGATCCGAGCAATCTCAGGCACTAAAGGATCATCTGATTGGG 337
 XX
 QY 63 ATCGACGCTCAGAGACACCGCGG 85
 Db 336 ATCACACACAGAGAACAGATGG 314
 XX
 RESULT 9
 AAT79814/c
 ID AAT79814 standard; cDNA; 444 BP.
 XX
 AC AAT79814;
 XX
 DT 12-NOV-1997 (first entry)
 XX
 DE cDNA encoding human ubiquitin conjugating enzyme 7.
 XX
 KW Ubiquitin conjugating enzyme; UCE; lymphocyte homing receptor;

KW growth hormone; viral infection; apoptosis inhibition; platelet;
 KW programmed cell death; AIDS; acquired immune deficiency syndrome;
 KW immunological disorder; Alzheimer's disease; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..444
 FT /*tag= a
 XX
 PN US650313-A.
 XX
 PD 22-JUL-1997.
 XX
 PF 05-JUN-1995; 95US-0464342.
 XX
 PR 05-JUN-1995; 95US-0464342.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Adams MD, Gentz R, Ni J;
 XX
 DR WPI; 1997-384673/35.
 DR P-PSDB; AAW5114.
 XX
 PT Ubiquitin conjugating enzyme-7, -8 and -9 and encoding DNA - used in
 PT the treatment of diseases characterised by unprogrammed cell death,
 PT and viral infection
 XX
 PS Example 1; Column 29-30; 32pp; English.
 XX
 CC AAT79814 is a cDNA sequence encoding human ubiquitin conjugating enzyme
 CC (UCE) 7. UCE polypeptides were expressed intracellularly via gene
 CC therapy and can be used to provide a signal for the lymphocyte
 CC homing receptor, thereby regulating lymphocyte trafficking. The growth
 CC hormone receptor is modified by ubiquitin and UCE polypeptides may be
 CC used to regulate growth hormone activity. UCE polypeptides can be used
 CC to overcome viral infection by suppressing programmed cell death, e.g.
 CC in the treatment of AIDS. They can also be used to inhibit the cytotoxic
 CC properties of platelets and the production of oxygen metabolites by
 CC platelets. They may be used to regulate immunological disorders in which
 CC platelets seem to be involved, e.g. hymenoptera venom hypersensitivity
 CC and aspirin-sensitive asthma. The UCE polypeptides may also be used to
 CC treat malignant transformation because proto-oncoproteins c-Mos and
 CC v-Jun are degraded in a ubiquitin-dependent manner. UCE antagonists,
 CC selected from UCE mutants, antisense DNA or smaller molecules which
 CC are able to pass through the cell membrane and bind to the proteins
 CC catalytic sites, may be used in the treatment of diseases in which UCE
 CC catalyses the transfer of ubiquitin to a substrate and marks that
 CC substrate for cell death, e.g. Alzheimer's disease and endemic
 CC pemphigus foliaceus.
 XX
 SQ Sequence 444 BP; 135 A; 95 C; 91 G; 123 T; 0 other;
 XX
 Query Match 17.6%; Score 30.2; DB 18; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 XX
 QY 3 TTCTCATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAGAATCGAATTGAG 62
 Db 396 TTCTCTATCTGTTTGTGATCCGAGCAATCTCAGGCACTAAAGGATCATCTGATTGGG 337
 XX
 QY 63 ATCGACGCTCAGAGACACCGCGG 85
 Db 336 ATCACACACAGAGAACAGATGG 314
 XX
 RESULT 10
 AA225304/c
 ID AA225304 standard; cDNA; 444 BP.
 XX
 AC AA225304;
 XX

DT 16-DEC-1999 (first entry)
 XX Human ubiquitin conjugating enzyme 7 encoding cDNA.
 DE
 XX
 KM Human; ubiquitin conjugating enzyme; UBE7, UBE8, UBE9; proliferation;
 KM Alzheimer's disease; atrophying skeletal muscle; diagnosis;
 KM African swine fever; apoptotic cell death; detection;
 KM cervical carcinoma; endemic pemphigus foliaceus; immunological disorder;
 KM malignant transformations; ss.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT CDS 1..444
 FT /*tag= a
 XX
 PN US968797-A.
 XX
 PD 19-OCT-1999.
 XX
 PF 22-JUL-1997; 97US-0903396.
 XX
 PR 05-JUN-1995; 95US-0464342.
 PR 31-JAN-1995; 95WO-US01250.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Adams MD, Gentz R, NI J;
 XX
 DR WPI, 1999-590409/50.
 DR P-PSDB; AAY42150.
 XX
 PT New human ubiquitin conjugating enzymes useful for treating malignant
 PT transformations and immunological disorders -
 PS
 PS Example 1; Fig 1; 33pp; English.
 XX
 CC The present sequence encodes human ubiquitin conjugating enzyme (UCE) 7.
 CC UBE enzymes and the DNA encoding them are useful for treating malignant
 CC transformations, immunological disorders, marking unwanted cells for
 CC death and to screen for agonists and antagonists which interact with
 CC the polypeptides. UBE antagonists can be used in the treatment of
 CC atrophying skeletal muscle, cervical carcinoma, Alzheimer's disease,
 CC endemic pemphigus foliaceus and African swine fever.
 CC
 XX
 SQ Sequence 444 BP; 135 A; 95 C; 91 G; 123 T; 0 other;
 XX
 Query Match 17.6%; Score 30.2; DB 20; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 3 TTTCATATCCGGAGAGGACAGAGCTATTTCAGCCATGAAAGCATCGAATTGAG 62
 DB 396 TTCTCTATCTGTTTGTGATCCGAGCAATCTCAGGACCTAAAGATCATCTGATTGGG 337
 QY 63 ATCGCAGCTCAGAGACACCGGG 85
 DB 336 ATCACACACAGAGAACAGATGG 314
 XX
 RESULT 11
 AA27558/c
 ID AA27558 standard; cDNA; 444 BP.
 XX
 AC AA27558;
 XX
 DT 15-DEC-1999 (first entry)
 XX
 DE Human ubiquitin conjugating enzyme coding sequence.
 KM Ubiquitin conjugating enzyme; UBE; ubiquitin-mediated proteolysis;
 KM cell-cycle regulatory protein; ubiquitination inhibitor; atherosclerosis;
 KM proliferative disorder; cancer; restenosis; tissue connective disorder;

KM wound healing; fibrosis disorder; rheumatoid arthritis; scleroderma;
 KM insulin dependent diabetes mellitus; glomerulonephritis; cirrhosis;
 KM diagnosis; therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN US968761-A.
 XX
 PD 19-OCT-1999.
 XX
 PF 07-JUN-1995; 95US-0486663.
 XX
 PR 04-JAN-1994; 94US-0176937.
 PR 23-MAY-1994; 94US-0247904.
 PR 27-MAY-1994; 94US-0250795.
 PR 13-SEP-1994; 94US-0305520.
 XX
 PA (MITO-) MITOTIX INC.
 XX
 PI Chiu MI, Cottarel G, Berlin V, Damagnez V, Draetta G, Rolfe M;
 XX
 DR WPI, 1999-590402/50.
 DR P-PSDB; AAY39961.
 XX
 PT Identifying ubiquitination inhibitors using novel ubiquitin conjugating
 PT enzymes -
 PS
 PS Claim 1; Column 67-70; 61pp; English.
 XX
 CC This sequence encodes a ubiquitin conjugating enzyme (UBCE). The
 CC invention relates to assays for identifying an inhibitor of
 CC ubiquitin-mediated proteolysis of a cell-cycle regulatory protein
 CC comprising contacting a candidate agent with an ubiquitin-conjugating
 CC system and measuring the level of ubiquitination. The
 CC ubiquitin-conjugating system comprises: (a) a reconstructed protein
 CC mixture including a ubiquitin conjugating enzyme (UBCE) produced by the
 CC expression of a nucleic acid which hybridizes under high stringency
 CC conditions to human UBCE, Candida albicans UBCE, or Schistosomaworms
 CC pombe UBCE coding sequences; (b) a regulatory protein; and (c) ubiquitin.
 CC The polynucleotides are useful for identifying ubiquitination inhibitors.
 CC The polynucleotides, polypeptides, antisense compounds and antibodies
 CC against them may also be useful for the treatment and/or diagnosis of
 CC proliferative disorders (e.g. cancer, atherosclerosis, or restenosis),
 CC tissue connective disorders, controlling wound healing, and disorders
 CC characterized by fibrosis (e.g. rheumatoid arthritis, insulin dependent
 CC diabetes mellitus, glomerulonephritis, cirrhosis, and scleroderma).
 CC
 XX
 SQ Sequence 444 BP; 132 A; 99 C; 91 G; 122 T; 0 other;
 XX
 Query Match 17.6%; Score 30.2; DB 20; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 3 TTTCATATCCGGAGAGGACAGAGCTATTTCAGCCATGAAAGCATCGAATTGAG 62
 DB 396 TTCTCTATCTGTTTGTGATCCGAGCAATCTCAGGACCTAAAGATCATCTGATTGGG 337
 QY 63 ATCGCAGCTCAGAGACACCGGG 85
 DB 336 ATCACACACAGAGAACAGATGG 314
 XX
 RESULT 12
 AAX78496/c
 ID AAX78496 standard; cDNA; 444 BP.
 XX
 AC AAX78496;
 XX
 DT 09-SEP-1999 (first entry)
 XX
 DE Human UBCE cDNA.
 KM Inhibitor; ubiquitin-mediated proteolysis; IkappaB polypeptide; HECT;

KW	ubiquitin-conjugating system; Homologous to E6-AP carboxyl terminus;
KW	Ilgase; ubiquitin; ubiquitination; RSC ligase; KLAN ligase; UBC4;
KM	MMD3; modulator; treatment; proliferative disorder; apoptosis; sepsis;
KM	differentiative disorder; viral infection; tissue wasting disorder;
KM	cachexia; malignancy; inflammatory disease; parasitic disease;
KM	tuberculosis; IL-2 therapy; rheumatoid arthritis; rheumatoid spondylitis;
KM	osteoarthritis; gouty arthritis; respiratory distress syndrome;
KM	pneumal malaria; chronic pulmonary inflammatory disease; silicosis;
KM	pulmonary sarcoidosis; bone resorption disease; reperfusion injury;
KM	graft versus host reaction; allograft rejection; Crohn's disease;
KM	systemic lupus erythematosus; multiple sclerosis; autoimmune diabetes;
KM	Homo sapiens.
OS	
XX	
XX	Key Location/Qualifiers
FH	1..444
FT	/+cag= a
FT	/product= "UBC4"
PN	MO9904033-A1.
PD	
XX	28-JAN-1999.
XX	
PF	16-JUL-1998; 98WO-US14638.
XX	
PR	16-JUL-1997; 97US-0895601.
PA	(MITO-) MITOTIX INC.
PI	Beer-Romero P, Glass SJ, Rolfe M, Strack PR;
XX	WPI; 1999-132274/11.
DR	P-PSDB; AAY25172.
PT	Identifying modulators of I-kappa-B proteolysis - used to develop
PT	products for treating e.g. proliferative and/or differentiative
PT	disorders, infections, tissue wasting, cachexia or AIDS
XX	
PS	Disclosure; Page 64-65; 79pp; English.
XX	
CC	This invention describes a novel assay identifying an inhibitor of
CC	ubiquitin-mediated proteolysis of a IkappaB polypeptide. The method
CC	comprises: (a) providing a ubiquitin-conjugating system including the
CC	IkappaB polypeptide and a HECT (Homologous to E6-AP Carboxyl Terminus)
CC	ligase and ubiquitin, to promote ubiquitination of the IkappaB
CC	polypeptide by the HECT ligase; (b) contacting the ubiquitin-conjugating
CC	system with a candidate agent; (c) measuring a level of ubiquitination of
CC	the polypeptide in the presence of the candidate agent and (d) comparing
CC	the measured level of ubiquitination in the presence of the candidate
CC	agent with ubiquitination of the IkappaB polypeptide in the absence of
CC	the candidate agent; where a statistically significant decrease in
CC	ubiquitination of the IkappaB polypeptide in the presence of the
CC	candidate agent is indicative of an inhibitor of ubiquitination of the
CC	IkappaB polypeptide. The assays can be used to identify compounds which
CC	modulate binding and/or ubiquitinylation of an IkappaB (or other cellular
CC	or viral substrate) by a HECT ligase, such as RSC or KLAN. Such
CC	modulators can be used e.g. in the treatment of proliferative and/or
CC	differentiative disorders, to modulate apoptosis, in the treatment of
CC	viral infections, and in the treatment of tissue wasting disorders e.g.
CC	cachexia secondary to infection or malignancy, cachexia secondary to
CC	human AIDS, inflammatory diseases, parasitic diseases, tuberculosis and
CC	high dose IL-2 therapy; rheumatoid arthritis, rheumatoid spondylitis,
CC	osteoarthritis, gouty arthritis and other arthritic conditions, sepsis,
CC	respiratory distress syndrome, cerebral malaria, chronic pulmonary
CC	inflammatory disease, silicosis, pulmonary sarcoidosis, bone resorption
CC	diseases, reperfusion injury, graft versus host reaction, allograft
CC	rejections, Crohn's disease, ulcerative colitis, or pyrexia, in addition
CC	to a number of autoimmune diseases such as multiple sclerosis, autoimmune
CC	diabetes, systemic lupus erythematosus, and ENL in leprosy, HIV, and
CC	AIDS. This sequence encodes a human UBC4 protein which is used in the
CC	method of the invention.
XX	

Sequence	444 BP; 132 A; 99 C; 91 G; 122 T; 0 other;
Query Match	17.6%; Score 30.2; DB 20; Length 444;
Best Local Similarity	60.2%; Pred. No. 1.2;
Matches	50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY	3 TTCCTAATACCGGGAGAGACACAGAGCTATTTCGCCCATGAAAGATCGGAATTGAG 62
DB	336 TTCCTATATCTGTTTGGTAGATCCAGCATCTCAGGACCTAAAGATCATCTGGATTGGG 337
OY	63 ATCGAGCTCAGAGACACCCGGG 85
DB	336 ATCAACAACAGAGAACGATGG 314
RESULT 13	
AAV82865/c	
ID	AAV82865 standard; cDNA; 444 BP.
AC	AAV82865;
XX	
DT	24-FEB-1999 (first entry)
XX	
DE	Ubiquitin conjugating enzyme (UCE) 7 cDNA.
XX	
XX	Human; Ubiquitin conjugating enzyme; UCE; UCE7; malignant transformation;
KW	c-Mos; v-Jun proto-oncoprotein; ubiquitin-dependent degradation;
KW	immunological disorder; hypersensitivity; aspirin-sensitive asthma;
KW	lymphocyte trafficking; growth receptor; viral infection;
KW	acquired immune deficiency syndrome; skeletal muscle atrophy;
KW	cervical cancer; endemic pemphigus foliaceus; African swine fever;
KW	cell death; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS 1..444
FT	/*tag= a
FT	/product= UCE7
XX	
PN	US5849286-A.
PD	
XX	15-DEC-1998.
XX	
PF	05-JUN-1995; 95US-0464604.
XX	
PR	05-JUN-1995; 95US-0464604.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
P1	Adams MD, Gentz R, N1 J;
XX	
DR	WPI: 1999-069655/06.
XX	P-PSDB; AAM85561.
XX	
PT	Human ubiquitin conjugating enzymes and related nucleic acid -
PT	useful for treating or diagnosing e.g. cancer, skeletal muscle
PT	atrophy, immunological diseases and infections
XX	
PS	Example 1; Fig 1A-B; 34pp; English.
XX	
CC	The present sequence encodes a human ubiquitin conjugating enzyme 7
CC	(UCE7). The UCE products are used to treat malignant transformation
CC	(where associated with c-Mos and v-Jun proto-oncoproteins, since these
CC	undergo ubiquitin-dependent degradation) or immunological disorders
CC	(e.g. hypersensitivity to wasp/bee stings or aspirin-sensitive asthma),
CC	to mark cells (particularly virus-infected) for death and to screen for
CC	agents (antagonists and agonists) that interact with them. The UCE
CC	products are also used to regulate lymphocyte trafficking or activity of
CC	the growth receptor, to treat many viral infections (by overcoming
CC	virus-induced suppression of apoptosis), for treating immune suppression
CC	(particularly acquired immune deficiency syndrome), to treat cytotoxic
CC	effects of platelets and their production of oxygenated metabolites. The

CC UCE protein is used to raise antibodies, useful in diagnostic
 CC immunosays and as a therapeutic inhibitor. Antagonists of the UCE
 CC protein are used to treat atrophy of skeletal muscle, cervical cancer
 CC (and some other tumors), endemic pemphigus foliaceus and African swine
 CC fever, or generally any condition in which UCE catalyses transfer of
 CC ubiquitin to a substrate, marking it for cell death.

XX Sequence 444 BP; 135 A; 95 C; 91 G; 123 T; 0 other;

Query Match 17.6%; Score 30.2; DB 20; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGCTATTTCAGCCACATGAAAGCATCGGAATTGAG 62
 DB 336 TTCTCATCTGTTTGTGATCCGAGCAATCTCAAGCATCTGATTCGATTGGG 337
 QY 63 ATCGAGCTCAGAGACACCCGGG 85
 DB 336 ATCACAACAAGAAACAGATGG 314

RESULT 14
 ID AAA61613/c
 XX AAA61613 standard; cDNA; 444 BP.

AC AAA61613;
 DT 23-OCT-2000 (first entry)

DE cDNA encoding human ubiquitin-conjugating enzyme hUBC.

XX hUBC; ubiquitin conjugating enzyme; ubiquitin mediated proteolysis;
 KW human; cellular protein half life; ubiquitination inhibitor; p53;
 KW cyclin; cell cycle regulator; myc deregulation; human papillomavirus;
 KW HPV-18 E6 protein; cervical cancer; skin cancer; epidermal hyperplasia;
 KW epidermal neoplasia; psoriasis; connective tissue disorder;
 KW wound healing; cytostatic; antiproliferative; anticancer; antipsoriatic;
 KW ss.

OS Homo sapiens.

XX Key Location/Qualifiers
 FT 1..444
 FT /*tag= a
 FT /product= "hUBC"

XX US6068982-A.

XX 30-MAY-2000.

XX 17-DEC-1996; 96US-0767942.

XX 07-JUN-1995; 95US-0486663.

XX 04-JAN-1994; 94US-0176937.

XX 23-MAY-1994; 94US-0247904.

XX 27-MAY-1994; 94US-0250795.

XX 13-SEP-1994; 94US-0305520.

XX (MITO-) MITOTIX INC.

XX Chiu MI, Cottarel G, Berlin V, Draetta G, Damagnez V, Rolfe M;

XX WPI, 2000-410854/35.

XX P-PSDB; AAB03169.

XX Identifying an inhibitor of ubiquitin mediated proteolysis of
 PT regulatory protein for treating cancers involves measuring
 PT ubiquitination levels of the protein in the presence of candidate agent
 PT in an eukaryotic cell
 XX
 XX Claim 1; Column 67-70; 73pp; English.

CC The invention relates to a method of identifying an inhibitor of
 CC ubiquitin-mediated proteolysis of a cell cycle regulatory protein
 CC comprising contacting an engineered eukaryotic cell with a candidate
 CC agent. The eukaryotic cells is engineered to express a recombinant
 CC human, Candida albicans or Schizosaccharomyces pombe ubiquitin-
 CC conjugating enzyme (AAB03169-B03171), a cell cycle regulatory protein
 CC (such as p53) and ubiquitin. The specification also discloses novel
 CC Candida albicans and Schizosaccharomyces pombe ubiquitin-conjugating
 CC enzymes, cadDC8 and spUBC (AAB03170, AAB03171), and two novel human
 CC ubiquitin-conjugating enzymes, hUBC and rapUBC (AAB03169, AAB03173).
 CC The ubiquitin-mediated proteolysis system is the major pathway for the
 CC selective, controlled degradation of intracellular proteins in
 CC eukaryotic cells. In particular, this system controls the half-lives of
 CC cellular proteins, and is important in controlling the levels of proteins
 CC involved in cell cycle progression. Alterations in the ubiquitination of
 CC these proteins may therefore play a role in the development of cancers.
 CC For example, human papillomavirus such as HPV-18 encode a transforming
 CC protein, E6 (AAB03176), which combines with a cellular E6-associated
 CC protein (E6-AP; AAB03177) to stimulate the ubiquitination of p53, thus
 CC targeting it for degradation. The ubiquitination inhibitors identified
 CC according to the method of the invention are useful for treatment of
 CC cervical cancers and connective tissue disorders and for controlling the
 CC wound healing process. They are also useful in treatment of hyperplastic
 CC epidermal conditions such as psoriasis, neoplastic epidermal conditions,
 CC skin cancers e.g., basal cell carcinomas, squamous cell carcinomas. The
 CC inhibitors are useful for deregulating myc expression and rendering the
 CC cells sensitive to chemotherapeutic treatment or to upset the balance of
 CC transformed cells and cause apoptosis to occur. Inhibitors of ubiquitin-
 CC mediated degradation of cyclins are useful as antiproliferative agents.
 CC The present sequence represents cDNA encoding the human ubiquitin-
 CC conjugating enzyme hUBC.

XX Sequence 444 BP; 132 A; 99 C; 91 G; 122 T; 0 other;

Query Match 17.6%; Score 30.2; DB 21; Length 444;
 Best Local Similarity 60.2%; Pred. No. 1.2;
 Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCAATACCGGAGAGGACAGCTATTTCAGCCACATGAAAGCATCGGAATTGAG 62
 DB 336 TTCTCATCTGTTTGTGATCCGAGCAATCTCAAGCATCTGATTCGATTGGG 337
 QY 63 ATCGAGCTCAGAGACACCCGGG 85
 DB 336 ATCACAACAAGAAACAGATGG 314

RESULT 15

AAAT41769/c
 ID AAAT41769 standard; DNA; 452 BP.

XX AAAT41769;

XX 24-JAN-1997 (first entry)

XX Human ubiquitin-conjugating enzyme UB4 coding sequence.

XX Ubiquitin-conjugating enzyme; UB4; cell cycle regulatory protein;
 KW CCRP; inhibitor; apoptosis; cell differentiation; ss.

XX Homo sapiens.

XX WO9633286-A1.

XX 24-OCT-1996.

XX 19-APR-1996; 96WO-US05643.

XX 20-APR-1995; 95US-0425299.

XX (HARD) HARVARD COLLEGE.

XX King RW, Kirschner MW, Peters J;

XX WPI; 1996-485790/48.
DR P-PSDB; AAM00363.
XX

PT Detecting inhibitors of ubiquitin-mediated proteolysis of CCRPs -
PT used in the treatment of proliferative and/or differentiation
PT diseases, and in modulation of apoptosis

XX
PS Example 1; Page 52; 63pp; English.

XX
CC A cDNA sequence (AAT41769) comprises the coding region for human
CC ubiquitin-conjugating enzyme UBC4 (AAM00362). It was obtd. from
CC an HeLa cDNA library by PCR amplification (see also AAT41765-68).
CC The cDNA was incorporated into a baculovirus vector and used to
CC produce recombinant UBC4 in Spodoptera frugiperda Sf9 cells, and
CC into a pGEX vector to allow UBC4 prodn. in E. coli. UBC4 can be
CC used in a novel ubiquitin-conjugating system to identify inhibitors
CC of ubiquitin-mediated proteolysis of cell cycle regulatory proteins.

XX
SQ Sequence 452 BP; 134 A; 101 C; 92 G; 125 T; 0 other;

Query Match 17.6%; Score 30.2; DB 17; Length 452;
Best Local Similarity 60.2%; Pred. No. 1.3;
Matches 50; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 3 TTCTCATACCGGAGAGGCAAGACTATTTCAGCCCATGAAAGCATCGAATTGAG 62
DB 405 TTCTCATACCGGAGAGGCAAGACTATTTCAGCCCATGAAAGCATCGAATTGAG 346
QY 63 ATCGAGCTCAGAGGACACCGGG 85
DB 345 ATCAACACACAGAGACAGATGG 323

Search completed: April 15, 2003, 18:58:06
Job time : 65.8981 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:52:31 ; Search time 356.738 Seconds

(without alignments)
7808.593 Million cell updates/sec

Title: US-09-647-019-6

Perfect score: 172
Sequence: 1 ggttcacataccgggagag.....taaacatacataaagcag 172

Scoring table: IDENTITY NUC
Gapop 10-0, Gapexc 1.0

Searched: 1615406 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estdb: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_estro: *
9: em_estro: *
10: gb_est1: *
11: gb_est2: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_g88: *
18: em_g88_hum: *
19: em_g88_inv: *
20: em_g88_pln: *
21: em_g88_vtc: *
22: em_g88_fun: *
23: em_g88_mam: *
24: em_g88_mus: *
25: em_g88_other: *
26: em_g88_pro: *
27: em_g88_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	172	100.0	280	AA248485	AA248485 cbh0287.b
2	160	93.0	909	BF790243	BF790243 602249777
3	152	88.4	277	N86813	N86813 L1050P Huma
4	152	88.4	320	W07478	W07478 za96c10.r1
5	151	87.8	360	A1355905	A1355905 3055-10R-
6	136	79.1	867	BF790636	BF790636 602250354

7	133	77.3	756	12	BF791178	BF791178 602251278
8	133	77.3	793	12	BF671987	BF671987 602152408
9	127	73.8	719	14	BM697544	BM697544 UT-R-DXO-
10	119.4	69.4	871	12	BF693124	BF693124 602080151
11	113.2	65.8	263	9	AA247872	AA247872 j3204.fseq
12	113.2	65.8	263	14	N55839	N55839 J3204F Huma
13	109	63.4	394	9	AA389647	AA389647 M104 Feta
14	106.4	61.9	542	12	BF575112	BF575112 602134792
15	104.2	60.6	517	12	BF395174	BF395174 UR-R-CMO-
16	104.2	60.5	517	12	BG224245	BG224245 LM0028P05
17	104	60.5	159	9	AA248067	AA248067 cp1451.9e
18	100	58.1	906	12	BF672126	BF672126 602152580
19	98	57.0	450	9	AA214147	AA214147 zn58f02.r
20	97	56.4	319	14	N87511	N87511 L1119F Hum
21	97	56.4	459	9	AA800221	AA800221 EST189718
22	95.6	55.6	491	12	BF284896	BF284896 EST449487
23	95.2	55.3	330	9	AA763276	AA763276 v099d04.r
24	95.2	55.3	364	14	W97451	W97451 m197c02.r1
25	95.2	55.3	368	14	W29186	W29186 mc22f02.r1
26	95.2	55.3	448	14	W13738	W13738 mb32a12.r1
27	95.2	55.3	482	9	AA434782	AA434782 ve23c01.r
28	95.2	55.3	934	11	AK003105	AK003105 Mue muscu
29	94.6	55.0	507	12	BF395391	BF395391 UT-R-CMO-
30	94.4	54.7	934	11	AK010172	AK010172 Mue muscu
31	94	54.7	486	9	A1035961	A1035961 ub50b12.r
32	94	54.7	510	10	AA918749	AA918749 B8T350053
33	94	54.7	613	14	B0554133	B0554133 H402SC08-
34	93.6	54.4	540	11	AK009857	AK009857 M08 muscu
35	91	52.9	450	9	AA800829	AA800829 EST190326
36	89.2	51.9	303	10	BB565182	BB565182 B8565182
37	89	51.7	416	9	A1153970	A1153970 ud49h02.r
38	89	51.7	1031	12	BF693607	BF693607 602081956
39	88	51.2	855	12	BF672902	BF672902 602152759
40	87	50.6	214	9	AA092554	AA092554 J15591.be
41	86.4	50.2	225	9	AA249531	AA249531 J15022.be
42	86.4	50.2	225	14	N56276	N56276 J15022P Hum
43	85	49.4	480	9	AA214155	AA214155 zn58f10.r
44	85	49.4	501	9	AA211521	AA211521 zn55b01.r
45	84	48.8	504	9	AA033164	AA033164 m137c11.r

ALIGNMENTS

RESULT 1
LOCUS AA248485 280 bp mRNA linear EST 11-MAR-1997
DEFINITION cbh0287.seq, F Human fetal heart, lambda ZAP Express Homo sapiens
ACCESSION AA248485
VERSION AA248485.1 GI:1879506
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (base 1 to 280)
Liew, C.C.
TITLE CDNA from human fetal heart (1997)
JOURNAL Unpublished (1997)
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: clliew@rics.bwh.harvard.edu
PCR PRIMERs
FORWARD: 5' GCCAGCTCGAATTAACTCACTCAATAAGG 3'
REVERSE: 5' CCACTGATTTGATATACGACTACTATAGGGCG 3'
Seq primer: 5' GAATTAACTCACTCAATAAGG 3'.
Location/Qualifiers
1. 280

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="Human fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested Lambda ZAP Express."

BASE COUNT 78 a 74 c 68 g 60 t

Query Match 100.0%; Score 172; DB 9; Length 280;
Best Local Similarity 100.0%; Pred. No. 2,9e-45;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCTCAATACCGGAGAGACAGAGCTATTTCAGCCATGAAAGCATCGGAATTG 60
DB 1 GGTCTCAATACCGGAGAGACAGAGCTATTTCAGCCATGAAAGCATCGGAATTG 60

QY 61 AGATCGAGCTCAGAGACACCGGGGCGCCCTTCCACCTTCCAGAGCTTTGATCTT 120
DB 61 AGATCGAGCTCAGAGACACCGGGGCGCCCTTCCACCTTCCAGAGCTTTGATCTT 120

QY 121 GCATCTGGCTGCTGGGACTTCCCTTAGGAGTAACAATATCATTAAGCAG 172
DB 121 GCATCTGGCTGCTGGGACTTCCCTTAGGAGTAACAATATCATTAAGCAG 172

RESULT 2
BF790243 909 bp mRNA linear EST 12-JAN-2001
DEFINITION 602249777F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
mRNA sequence.
ACCESSION BF790243
VERSION BF790243.1 GI:12095188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 909)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/MLML at: http://image.llnl.gov
Plate: LICM1187 row: 1 column: 19
High quality sequence stop: 626.
Location/Qualifiers
1. .909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328154"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgcccggcc); Site 2: SfiI (ggccattggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTAATGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGGAGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length

clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 279 a 200 c 212 g 218 t

Query Match 93.0%; Score 160; DB 12; Length 909;
Best Local Similarity 99.4%; Pred. No. 2,7e-41;
Matches 171; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GGTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAGCATCGGAATTG 60
DB 1 GGTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAGCATCGGAATTG 60

QY 61 AGATCGAGCTCAGAGACACCGGGGCGCCCTTCCACCTTCCAGAGCTTTGATCTT 120
DB 61 AGATCGAGCTCAGAGACACCGGGGCGCCCTTCCACCTTCCAGAGCTTTGATCTT 119

QY 121 GCATCTGGCTGCTGGGACTTCCCTTAGGAGTAACAATATCATTAAGCAG 172
DB 120 GCATCTGGCTGCTGGGACTTCCCTTAGGAGTAACAATATCATTAAGCAG 171

RESULT 3
N86813 277 bp mRNA linear EST 01-APR-1996
LOCUS N86813
DEFINITION L1050F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
clone L1050 5' similar to EST (F9260), mRNA sequence.
ACCESSION N86813
VERSION N86813.1 GI:1440015
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 277)
Liew,C.C.
CDNA from fetal heart (1996)
Unpublished (1996)
Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliw@rics.bwh.harvard.edu
Seq primer: GAATTAACCTCCTCACTAAGG.
Location/Qualifiers
1. .277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="L1050"
/clone_lib="Human fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested Lambda ZAP Express."

BASE COUNT 79 a 71 c 62 g 65 t

Query Match 88.4%; Score 152; DB 14; Length 277;
Best Local Similarity 100.0%; Pred. No. 9,2e-39;
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 GCAGAGGCTATTTCAGCCATGAAAGCATGGAATTGAGATCGAGCTCAGAGACA 80
DB 1 GCAGAGGCTATTTCAGCCATGAAAGCATGGAATTGAGATCGAGCTCAGAGACA 60

QY 81 CCGGGGCGCCCTTCCACCTTCCAGAGAGCTTTGATCTTGCATCTGGCTGGGACT 140
DB 61 CCGGGGCGCCCTTCCACCTTCCAGAGAGCTTTGATCTTGCATCTGGCTGGGACT 120

QY 141 TCCTTAGGAGTAAACAATACATAAGCAG 172
 DB 121 TCCTTAGGAGTAAACAATACATAAGCAG 152

RESULT 4
 LOCUS W07478 320 bp mRNA linear EST 25-Apr-1996
 DEFINITION 2896c10.r1 Soares fetal lung NBHL19W Homo sapiens cDNA clone
 IMAGE:300402 5', mRNA sequence.
 W07478
 VERSION W07478.1 GI:1281479
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 320)
 HILLIER, L., CLARK, N., DUBUQUE, T., ELLISTON, K., HAWKINS, M., HOLMAN, M., HULTEMAN, M., KUCABA, T., LE, M., LEMMON, G., MARRA, M., PARSONS, J., RIFKIN, L., ROHLFING, T., SOARES, M., TAN, F., TREVAEKIE, E., WATERSTON, R., WILLIAMSON, A., WOHLDAMM, P. and WILSON, R.
 The Wash-Merck EST Project
 Unpublished (1995)
 CONTACT: Wilson R.
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: mob.REGA+ET
 High quality sequence stop: 279.

FEATURES

source

1..320
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="GDB:1245326"
 /db_xref="taxon:9606"
 /clone="IMAGE:300402"
 /clone_1lb="Soares_fetal_lung_NBHL19W"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: lung; Vector: pTZ19 (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo (dT) primer [5'-GTGTACCAATCTGAGAGTGGAGGCGGCAATTTTGTGTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Patricia Bonaldi. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBHL19W."

BASE COUNT 92 a 83 c 77 g 67 t 1 others
 ORIGIN

Query Match 88.4%; Score 152; DB 14; Length 320;
 Best Local Similarity 100.0%; Pred. No. 9.4e-39;
 Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 GCACAGAGCTATTTCAGCAGCATGAAAGCATGGAAATTGATCGAGCTCAGAGACA 80
 DB 2 GCACAGAGCTATTTCAGCAGCATGAAAGCATGGAAATTGATCGAGCTCAGAGACA 61
 QY 81 CCGGGGCCCCCTTCACCTTCAGAGAGCTTTGTATCTTGATCTGCTGCTGGAGCT 140
 DB 62 CCGGGGCCCCCTTCACCTTCAGAGAGCTTTGTATCTTGATCTGCTGCTGGAGCT 121
 QY 141 TCCTTAGGAGTAAACAATACATAAGCAG 172

DB 122 TCCTTAGGAGTAAACAATACATAAGCAG 153

RESULT 5
 LOCUS A1355905 360 bp mRNA linear EST 31-Dec-1999
 DEFINITION 3055-10R-8H11-2 Human heart cDNA (CCLee) Homo sapiens cDNA clone
 xh8H11.3', mRNA sequence.
 A1355905
 VERSION A1355905.1 GI:6649247
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 360)
 Zhuchenko, O., Patzak, D. and Wehnert, M.
 ESTs of clone xh8H11 (human heart)
 Unpublished (1999)
 Other ESTs: 8H11-R1-C8-OZ, 8H11-F1-C7-OZ, 3055-10U-8H11-2
 Department of Molecular Human Genetics
 Contact: Patzak D., submitter; (Wehnert M., supervisor)
 Institut für Human Genetik
 Fleischmannstr. 42/44, D-17487 Greifswald, Germany
 Tel: +49 3834 8653-78 (-74)
 Fax: +49 3834 8653-93
 Email: patzak@z.uni-greifswald.de (mwehnert@z.uni-greifswald.de)

Single path sequence (manual checked and edited) with a vector primer of clone xh8H11. Sequence starts with base 1 following the determined vector sequence (ACCGCGTGGCGCGCT, small letters: NotI-site). ESTs 3055-10R-8H11-2, 3055-10U-8H11-2 (with vector primers) and 8H11-R1-C8-OZ, 8H11-F1-C7-OZ (with gene specific primers) provide a contig of the whole insert of clone xh8H11.
 Insert Length: 850 Std Error: 10.00
 Plate: 8 row: H column: 11
 Seq primer: 3055-10R (like M13 reverse)
 High quality sequence stop: 360.

FEATURES

source

1..360
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="xh8H11"
 /clone_1lb="Human heart cDNA (CCLee)"
 /sex="male"
 /tissue_type="heart muscle"
 /dev_stage="adult"
 /lab_host="E. coli"
 /note="Organ: heart; Vector: Lambda ZAP II, PKSII-; Site 1: NotI; Site 2: NotI; Human heart cDNA library was constructed in Lambda ZAP II vectors using NotI linkers. Clones from the primary cDNA library were deposited into 96-well trays for storage and retrieval. The 'Isolation of chromosome-specific genes by reciprocal probing of arrayed cDNA and cosmid libraries' (Human Molecular Genetics, 1995 Vol. 4, No 8: p1373-1380) provided cDNA-clones as plasmids (vector PKSII-, E. coli)."

BASE COUNT 109 a 92 c 88 g 71 t
 ORIGIN

Query Match 87.8%; Score 151; DB 9; Length 360;
 Best Local Similarity 100.0%; Pred. No. 2e-38;
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 CACAGAGCTATTTCAGCAGCATGAAAGCATGGAAATTGATCGAGCTCAGAGACA 81
 DB 1 CACAGAGCTATTTCAGCAGCATGAAAGCATGGAAATTGATCGAGCTCAGAGACA 60
 QY 82 CCGGGGCCCCCTTCACCTTCAGAGAGCTTTGTATCTTGATCTGCTGCTGGAGCT 141
 DB 61 CCGGGGCCCCCTTCACCTTCAGAGAGCTTTGTATCTTGATCTGCTGCTGGAGCT 120
 QY 142 CCCTTAGGAGTAAACAATACATAAGCAG 172

REFERENCE 1 (bases 1 to 793)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 CDNA Library Preparation: CLONETECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLML at: <http://image.lnl.gov>
 Plate: LNCMI141 row: 0 column: 02
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES
 Source 1..793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4293721"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); site_1: SfiI (99ccgctcggcc); site_2: SfiI (ggcattacggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGGCGGCGGACATG-dr(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies inserted clones by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 251 a 164 c 175 g 203 t
 ORIGIN

Query Match 77.3%; Score 133; DB 12; Length 793;
 Best Local Similarity 100.0%; Pred. No. 1.6e-32;
 Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 ACATGAAAGCATCGAATTGAGATCGCAGCTCAGAGACACGGGGGCCCTCCACCT 99
 DB 2 ACATGAAAGCATCGAATTGAGATCGCAGCTCAGAGACACGGGGGCCCTCCACCT 61
 QY 100 TCAGAAGCTTTGATTTGATCTGCACTGCTGCTGCACTTCCCTTAGAGAGTAACAA 159
 DB 62 TCAGAAGCTTTGATTTGATCTGCACTGCTGCTGCACTTCCCTTAGAGAGTAACAA 121
 QY 160 ATACATTAAGCAG 172
 DB 122 ATACATTAAGCAG 134

RESULT 9 719 bp mRNA linear EST 28-FEB-2002
 LOCUS BM697544
 DEFINITION UI-E-DX0-agn-1-12-0-UI.r1 UI-E-DX0 Homo sapiens CDNA clone
 ACCESSION UI-E-DX0-agn-1-12-0-UI 5', mRNA sequence.
 VERSION BM697544.1 GI:19010802
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 719)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping

University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.reagen.com).
 Seq primer: M13 Reverse.

FEATURES
 Source 1..719
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-DX0-agn-1-12-0-UI"
 /clone_lib="UI-E-DX0"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-DX0 is a CDNA library containing the following tissue(s): fetal eyes. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AGAATCAGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 226 a 142 c 163 g 186 t
 ORIGIN
 Query Match 73.8%; Score 127; DB 14; Length 719;
 Best Local Similarity 100.0%; Pred. No. 1.4e-30;
 Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 AAAGCATCGAATTGAGATCGCAGCTCAGAGACACGGGGGCCCTCCACCTCCAG 105
 DB 1 AAAGCATCGAATTGAGATCGCAGCTCAGAGACACGGGGGCCCTCCACCTCCAG 60
 QY 106 GAGCTTTGATTTCTGATCTGCTGCTGCACTTCCCTTAGAGAGTAACAAATACAT 165
 DB 61 GAGCTTTGATTTCTGATCTGCTGCTGCACTTCCCTTAGAGAGTAACAAATACAT 120
 QY 166 AAGCAG 172
 DB 121 AAGCAG 127

RESULT 10 871 bp mRNA linear EST 22-DEC-2000
 LOCUS BF693124
 DEFINITION 602080151F1 NIH_MGC_81 Homo sapiens CDNA clone IMAGE:424364 5',
 ACCESSION BF693124
 VERSION BF693124.1 GI:11978532
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 871)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabds-remail.nih.gov
 Tissue Procurement: CLONETECH Laboratories, Inc.
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E.B. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 DNA distribution: MGC clone distribution information can be found through the I.M.A.G.E.B. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LLCM1056 row: f column: 13
 High quality sequence stop: 575.
 Location/Qualifiers

FEATURES

source

1. 871
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1db="IMAGE:424364"
 /lab_host="DH10B (TI phase-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgctcgcc); Site_2: SfiI (ggccatcagcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 280 a 157 c 218 g 216 t
 ORIGIN

Query Match

Best Local Similarity 69.4%; Score 119.4; DB 12; Length 871;
 Matches 131; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 40 ACATGAAAGCATCGAATTCGAGCTGAGAGACACCGGCGCCCTTCCACT 99

DB 2 ACATGAAAGCATCGAATTCGAGCTGAGAGACACCGGCGCCCTTCCACT 60

QY 100 TCAGAGAGCTTTGATTCCTGATCTGCTGCTGCTGCTGCTTGAAGAGTAACAA 159

DB 61 TCAGAGAGCTTTGATTCCTGATCTGCTGCTGCTGCTGCTTGAAGAGTAACAA 120

QY 160 ATACATAAGCAG 172

DB 121 ATACATAAGCAG 133

RESULT 11

AA247872

LOCUS 263 bp mRNA linear EST 11-MAR-1997

DEFINITION J3204.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA

ACCESSION AA247872

VERSION AA247872.1 GI:1880264

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 263)

AUTHORS Liew C.C.

JOURNAL Unpublished (1997)

COMMENT Contact: Liew CC

Brigham and Women's Hospital

Harvard Medical School

75 Francis St. Boston, MA 02115, USA

Tel: 6177328915

Fax: 6179750995

Email: cliew@rics.bwh.harvard.edu

PCR primers

FORWARD: 5' GCCAAGCTCGAATTAACCTCAGTAAGG 3'

BACKWARD: 5' CCAGTATGTGATAGCACTCAGTATGGGG 3'
 Seq primer: 5' GAATTAACCTCAGTAAGG 3'.
 Location/Qualifiers

FEATURES

source

1. 263
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1db="Human fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI, Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 76 a 71 c 58 g 58 t
 ORIGIN

Query Match

Best Local Similarity 65.8%; Score 113.2; DB 9; Length 263;
 Matches 138; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 28 GCTATTTACCCATGAAAGCATCGAATTCGAGCTCAGAGACACCGGCG 87

DB 1 GCTATTTACCCATGAAAGCATCGAATTCGAGCTCAGAGACACCGGCG 60

QY 88 CCCCTTCACCTTCAGAGAGCTTTGATTCCTGATCTGCTGCTGCTGCTGCT 145

DB 61 CCCCTTCACCTTCAGAGAGCTTTGATTCCTGATCTGCTGCTGCTGCTGCT 120

QY 146 TAGGACGTA-AACAAATACATAA 168

DB 121 TAGGACGTA-AACAAATACATAA 144

RESULT 12

N55839

LOCUS 263 bp mRNA linear EST 20-FEB-1996

DEFINITION J3204F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA

ACCESSION N55839

VERSION N55839.1 GI:1198687

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (bases 1 to 263)

AUTHORS Liew C.C.

JOURNAL Unpublished (1995)

COMMENT Contact: Liew CC

Brigham and Women's Hospital

Harvard Medical School

75 Francis St. Boston, MA 02115, USA

Tel: 6177328915

Fax: 6179750995

Email: cliew@rics.bwh.harvard.edu

Seq primer: TCCAGAGATTCGGCAGAG.

Location/Qualifiers

1. 263

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_1db="Human fetal heart, Lambda ZAP Express"

/lab_host="E. coli XL1-Blue"

/note="Vector: Lambda ZAP Express; Site 1: EcoRI, Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 76 a 71 c 58 g 58 t

ORIGIN

RESULT 15
 AF395174/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 BF395174
 UT-R-CMO-bj1-g-03-0-UI.s1
 UT-R-CMO-bj1-g-03-0-UI.3', mRNA sequence.
 BF395174
 BF395174.1 GI:11380139
 EST.
 542 bp
 RNA
 linear EST 27-NOV-2000
 notvegiclus cdna clone

SOURCE
ORGANISM

Norway rat
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
(Accession 1 to 5421)

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 542)	Bonaldo, M.F., Lennon, G. and Soares, M.B.	Normalization and subtraction; two approaches to facilitate gene discovery

JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msosares@blu.weeey.uwoa.edu
The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a nonide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the non-normalized heart library cDNA library preparation: M.B. Soares Lab Clones distribution: Clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 1-24, >AT-rich<low_complexity
Seq primer: M13 Forward
POLYA=yes.

FEATURES

Source

Location/Qualifiers

```

1..542
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db xref="taxon:10116"
/clone_uid="R-CMO-bj3-g-03-0-UI"
/clone_id="Ur-R-CMO"
/dev_stage="ADULT"
/lab_host="VDH0B (Life Technologies)"
/notes="Vector: pTR30-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The Ur-R-CMO
library is a non-normalized library constructed from rat
heart tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratcast.eng.uiowa.edu.
TAG LIB=Ur-R-CMO
TAG TISSUE=heart
TAG SEQ=ACMAC"

```

BASE COUNT	111 a	127 c	125 g	178 t	1 others
ORIGIN					

ORIGIN

Query Match	61.9%	Score 106.4;	DB 12;	Length 542;
Best Local Similarity	80.1%;	Pred. No. 6.8e-24;		
Matches 125; Conservative	0;	Mismatches 31;	Indels 0;	Gaps 0;

QY 17 AGAGGACAGAGCTATTTCACCCACGAAAGATGGAATTAGATCGAGCTCAG 76
 Db 530 AGAGCACTCCAGCTATTCCACCACATGAAAAAGACTGGAATTAGATCTCAGAG 411
 QY 77 GACACCGGGGCCCCCTTCACCTCCAGAGACTTGTATCTTGCACTTGCTGCTGG 136
 Db 470 GACACCGGAGAGTCTCTTACCCCTGTAAAGCGTTTCTGTTTTCACCTGGCTCTAG 411
 QY 137 GACTTCCCTTAGGCACTAAACAATATCATTAAGCAG 172
 Db 410 GACTGTCTCAGGCACTAAACAATCCAGAGAGCG 375

Search completed: April 15, 2003, 22:02:30
Job time : 359.738 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 / Search time 389.696 Seconds

(without alignments)
12645.099 Million cell updates/sec

Title: US-09-647-019-6

Perfect score: 172
Sequence: 1 ggtctcattacacgggagag.....taacaatacataaagcag 172

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sce:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sce:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_man:*
37: em_hcg_vtl:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	171	99.4	885	6	HS250584	AJ250584 Homo sapi
2	171	99.4	886	9	AX322774	AX322774 Sequence
3	171	99.4	886	9	AF129505	AF129505 Homo sapi
4	171	99.4	4740	9	U73509	U73509 Homo sapien
5	171	99.4	150319	2	AL772370	AL772370 Homo sapi
6	106.4	61.9	892	10	AF364071	AF364071 Rattus no
7	100	58.1	835	9	BC005948	BC005948 Homo sapi
8	95.2	55.3	787	10	AY026524	AY026524 Mus muscu
9	95.2	55.3	936	10	AF364070	AF364070 Mus muscu
10	95.2	55.3	943	10	MM0245772	MM0245772 Mus muscu
11	95.2	55.3	239667	2	AL731811	AL731811 Mus muscu
12	34.6	20.1	173333	9	AP002022	AP002022 Homo sapi
13	34.6	20.1	188818	2	AC098674	AC098674 Homo sapi
14	33.6	19.5	114594	9	AC110939	AC110939 Homo sapi
15	32.8	19.1	158785	9	AC079954	AC079954 Homo sapi
16	32.4	18.8	515	5	AB080118	AB080118 Oryzias l
17	32.4	18.8	54301	2	AC116199	AC116199 Rattus no
18	32.4	18.8	68222	9	AL391904	AL391904 Human DNA
19	32.4	18.8	148329	2	AC121327	AC121327 Oryza sat
20	32.4	18.8	176997	9	AC068953	AC068953 Homo sapi
21	32.4	18.8	177515	9	AP003468	AP003468 Homo sapi
22	32.4	18.8	194946	2	AC027362	AC027362 Homo sapi
23	32.2	18.7	266282	2	AC020963	AC020963 Mus muscu
24	32.2	18.6	108805	2	CNS08CBF	AL731875 Oryza sat
25	32	18.6	156825	2	AC026726	AC026726 Homo sapi
26	32	18.6	160666	2	AC068959	AC068959 Homo sapi
27	32	18.6	162069	2	AC108928	AC108928 Homo sapi
28	32	18.6	168247	2	AC073025	AC073025 Homo sapi
29	31.8	18.5	135245	2	AC116286	AC116286 Rattus no
30	31.8	18.5	160824	9	AC079228	AC079228 Homo sapi
31	31.6	18.4	1060	10	AF064717	AF064717 Mus spret
32	31.6	18.4	154621	2	AL133214	AL133214 Human DNA
33	31.6	18.4	174433	2	AC101909	AC101909 Mus muscu
34	31.6	18.4	176666	2	AC113090	AC113090 Mus muscu
35	31.6	18.4	191914	2	AC104671	AC104671 Mus muscu
36	31.2	18.1	84571	2	AC098375	AC098375 Rattus no
37	31.2	18.1	86574	9	HS83387	AL008637 Human DNA
38	31.2	18.1	120488	9	AC073257	AC073257 Homo sapi
39	31.2	18.1	190907	2	AC114732	AC114732 Homo sapi
40	31.2	18.1	243590	2	AC094989	AC094989 Rattus no
41	31	18.0	2015	10	BC003309	BC003309 Mus muscu
42	31	18.0	170367	9	AC012469	AC012469 Homo sapi
43	31	18.0	177835	9	AC087341	AC087341 Homo sapi
44	31	18.0	180745	9	AL359183	AL359183 Human DNA
45	31	18.0	186102	2	AC115802	AC115802 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS HS250584
DEFINITION Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
VERSION AJ250584.1 GI:10178976
KEYWORDS Strm gene, stretch responsive muscle (X-chromosome).
SOURCE human.
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 885)
Kemp,T.J., Sadusky,T.J., Simon,M., Brown,R., Baatwood,M.,
Sassoon,D.A. and Coulton,G.R.

TITLE Identification of a novel stretch-responsive skeletal muscle gene (smpx)

JOURNAL Genomics 72 (3), 260-271 (2001)

MEDLINE 21295047

PUBMED 11401441

REFERENCE 2 (bases 1 to 885)

AUTHORS Kemp, T.J.

TITLE Direct Submission

JOURNAL Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial College of Science Technology and Medicine, SAF Bldg, Level 2, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM

COMMENT Related sequences: AJ245772, U73508 to U73509.

FEATURES

source

1..885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="skeletal muscle"
1..885
/gene="Srmx"
1..183
/gene="Srmx"
184..450
/gene="Srmx"
/product="stretch responsive muscle (X-chromosome)"
/codon_start=1
/protein_id="CAC08482.1"
/db_xref="GI:10178977"
/translation="NMMSKQPSVNRATQAINIPMGAFRPGAGPPRRKCTPEVEE
GVPTSDERKRPFGAKLPGPAVNLSEIINKSELYVKRQ"

3'UTR
polyA_signal
857..862
/gene="Srmx"
857..862
/gene="Srmx"

BASE COUNT 278 a 172 c 191 g 244 t

ORIGIN

Query Match 99.4%; Score 171; DB 9; Length 885;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAAGCATCGAATTGA 61
DB 1 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAAGCATCGAATTGA 60

QY 62 GATCGAGCTCAGAGGACACCGGGCGCCCTTCCACCTTCCAGAGCTTTGATTCTTG 121
DB 61 GATCGAGCTCAGAGGACACCGGGCGCCCTTCCACCTTCCAGAGCTTTGATTCTTG 120

QY 122 CATCTGCTGCTGGGACTTCCCTTAGGAGTAACAATAATCATTAAGCAG 172
DB 121 CATCTGCTGCTGGGACTTCCCTTAGGAGTAACAATAATCATTAAGCAG 171

RESULT 2
AX322774 886 bp DNA linear PAT 07-JAN-2002
LOCUS AX322774
DEFINITION Sequence 18 from Patent WO192567.
ACCESSION AX322774
VERSION AX322774.1 GI:18093754
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 unidentified.
AUTHORS Bunk,D., Reuner,B., Beck,J. and Henkel,T.
TITLE Novel target genes for diseases of the heart
JOURNAL Patent: WO 0192567-A 18 06-DEC-2001;
Medigene AG (DE)
FEATURES
source Location/Qualifiers
1..886
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 278 a 172 c 191 g 245 t

ORIGIN

Query Match 99.4%; Score 171; DB 6; Length 886;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAAGCATCGAATTGA 61
DB 1 GTTCTCAATACCGGAGAGGACAGAGCTATTTCAGCCATGAAAAGCATCGAATTGA 60

QY 62 GATCGAGCTCAGAGGACACCGGGCGCCCTTCCACCTTCCAGAGCTTTGATTCTTG 121
DB 61 GATCGAGCTCAGAGGACACCGGGCGCCCTTCCACCTTCCAGAGCTTTGATTCTTG 120

QY 122 CATCTGCTGCTGGGACTTCCCTTAGGAGTAACAATAATCATTAAGCAG 172
DB 121 CATCTGCTGCTGGGACTTCCCTTAGGAGTAACAATAATCATTAAGCAG 171

RESULT 3
AF129505 886 bp mRNA linear PRI 22-DEC-1999
LOCUS AF129505
DEFINITION Homo sapiens small muscular protein (SMPX) mRNA, complete cds.
ACCESSION AF129505
VERSION AF129505.1 GI:6625646
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
AUTHORS Identification, mapping, and genomic structure of a novel X-chromosomal human gene (SMPX) encoding a small muscular protein Hum. Genet. 105 (5), 506-512 (1999)

JOURNAL MEDLINE 20065879
PUBMED 10598820
REFERENCE 2 (bases 1 to 886)
AUTHORS Patzak,D.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-1999) Molecular Human Genetics, Institut for Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany

FEATURES

source

1..886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22.1-22.2"
1..886
/gene="SMPX"
1..183
/gene="SMPX"
184..450
/gene="SMPX"
184..450
/gene="SMPX"
/codon_start=1
/evidence="not experimental"
/product="small muscular protein"
/protein_id="AAFI9343.1"
/db_xref="GI:6625647"
/translation="NMMSKQPSVNRATQAINIPMGAFRPGAGPPRRKCTPEVEE
GVPTSDERKRPFGAKLPGPAVNLSEIINKSELYVKRQ"

misc_feature 190
/gene="SMPX"
/note="alternate position for initiation methionine"
451..886
/gene="SMPX"
857..862
/gene="SMPX"

3'UTR
polyA_signal
857..862
/gene="SMPX"
172 c 191 g 244 t

BASE COUNT 279 a 172 c 191 g 244 t

ORIGIN

Query Match 99.4%; Score 171; DB 9; Length 886;
Best Local Similarity 100.0%; Pred. No. 2.1e-50;

Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTTCTCAATACCGGAGAGCAGAGCTATTTCAGCCACATGAAGAAGCATGGAAATTA 61

Db 1 GTTCTCAATACCGGAGAGCAGAGCTATTTCAGCCACATGAAGAAGCATGGAAATTA 60

Qy 62 GATCGAGCTCAGAGACACCGGCGCCCTTCACCTTCAGAGAGCTTTGATTTCTTG 121

Db 61 GATCGAGCTCAGAGACACCGGCGCCCTTCACCTTCAGAGAGCTTTGATTTCTTG 120

Qy 122 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 172

Db 121 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 171

RESULT 4
U73509/c 47440 bp DNA 1linear PRI 27-APR-1999
LOCUS Homo sapiens cosmid clone U228D4 from Xp22.1-22.2, complete
DEFINITION
ACCESSION U73509
VERSION U73509.1 GI:1616809
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 47440)
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998) ✓ November
MEDLINE 99063792

REFERENCE
AUTHORS 2 (bases 1 to 47440)
TITLE Waterston, R.
JOURNAL The sequence of Homo sapiens cosmid clone U228D4
REFERENCE 3 (bases 1 to 47440)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1996)
REFERENCE 4 (bases 1 to 47440)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-UTL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 47440)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sepiens@wustl.edu

COMMENT
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

FEATURES

source
1..47440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.1-22.2"
/clone="U228D4"
/clone_1ib="LLOXNCC01-U"
9..261
/rpt_family="Retroviral"
/rpt_family="MIR"
2020..2180
/rpt_family="MIR"
4824..4862
/rpt_family="(CAAT)n"
complement(5131..5222)
/note="match to EST N87511 (NID:g1440713)"
complement(5133..5222)
/note="match to EST W07478 (NID:g1281479) za96c10.r1"
complement(5148..5222)
/note="match to EST AA092554 (NID:g1637327)"
complement(5169..5222)
/note="match to EST AA248485 (NID:g1879506)"
7616..7763
/rpt_family="MIR"
8781..8842
/rpt_family="U2"
11721..11940
/rpt_family="MER1_type"
12770..12961
/rpt_family="MIR"
13039..13107
/rpt_family="L2"
13897..14201
/rpt_family="Alu"
14233..14397
/rpt_family="MIR"
complement(15622..15689)
/note="match to EST AA248485 (NID:g1879506)"
complement(15622..15689)
/note="match to EST W07478 (NID:g1281479) za96c10.r1"
16751..16837
/rpt_family="MIR"
complement(19314..19497)
/note="match to EST AA248485 (NID:g1879506)"
complement(19314..19478)
/note="match to EST W07478 (NID:g1281479) za96c10.r1"
complement(19314..19447)
/note="match to EST AA248067 (NID:g1878536)"
complement(19314..19423)
/note="match to EST N87511 (NID:g1440713)"
complement(19314..19413)
/note="match to EST AA092554 (NID:g1637327)"
21260..21572
/rpt_family="Retroviral"
21984..22282
/rpt_family="Alu"
23495..23673
/rpt_family="MIR"
24236..24556
/rpt_family="L1"
24759..24869

This clone was mapped by Grief, M., Whyte, M. P., Thakker, R. V.,
and Mazzarella, R. Sequence analysis of 139 kb in Xp22.1
containing spermine synthase and the 5' region of PEX. Genomics
44:227-231 (1997).

SOURCE INFORMATION:

This clone is from a chromosome X-specific cosmid library LLOXNCC01
'U'. The source of the chromosome was a human/hamster hybrid,
GM07297-F, from Robert Nussbaum at the University of Pennsylvania
School of Medicine. Please contact the Lawrence Livermore National
Laboratory at http://www-bio.llnl.gov/genome to obtain the clone.
VECTOR: Lawriett15

```

repeat_region      /rpt_family="L1"
25920. 26860
/rpt_family="L1"
repeat_region      /rpt_family="L1"
26867. 27210
/rpt_family="MALR"
repeat_region      /rpt_family="L1"
27213. 27810
/rpt_family="L1"
repeat_region      /rpt_family="Retroviral"
28727. 28816
/rpt_family="Retroviral"
repeat_region      /rpt_family="Retroviral"
28945. 29005
/rpt_family="Retroviral"
repeat_region      /rpt_family="L1"
29715. 30007
/rpt_family="L1"
repeat_region      /rpt_family="L1"
30773. 31151
/rpt_family="MALR"
repeat_region      /rpt_family="L1"
31705. 33575
/rpt_family="L1"
repeat_region      /rpt_family="L1"
33587. 37330
/rpt_family="L1"
repeat_region      /rpt_family="L1"
37345. 37464
/rpt_family="L1"
repeat_region      /rpt_family="L1"
37504. 37858
/rpt_family="L1"
repeat_region      /rpt_family="L1"
38102. 38146
/rpt_family="MALR"
repeat_region      /rpt_family="MALR"
38578. 38706
/rpt_family="L1"
repeat_region      /rpt_family="L1"
38794. 38914
/rpt_family="L1"
repeat_region      /rpt_family="Retroviral"
38939. 39240
/rpt_family="L1"
repeat_region      /rpt_family="L1"
39241. 39431
/rpt_family="Retroviral"
repeat_region      /rpt_family="Retroviral"
39469. 39659
/rpt_family="MER1_type"
repeat_region      /rpt_family="MER1_type"
39877. 39946
/rpt_family="L1"
repeat_region      /rpt_family="L1"
40027. 40328
/rpt_family="L1"
repeat_region      /rpt_family="L1"
42307. 42488
/rpt_family="MER1_type"
repeat_region      /rpt_family="MER1_type"
43876. 44349
/rpt_family="MALR"
repeat_region      /rpt_family="MALR"
45063. 45245
/rpt_family="MALR"
repeat_region      /rpt_family="MALR"
45276. 45399
/rpt_family="MER1_type"
repeat_region      /rpt_family="MER1_type"
46393. 46517
/rpt_family="L1"
repeat_region      /rpt_family="L1"
46645. 47185
/rpt_family="MER1_type"
repeat_region      /rpt_family="MER1_type"
46645. 47185
/rpt_family="MER1_type"
BASE COUNT      15150 a 9644 c 9310 g 13336 t
ORIGIN

```

```

Query Match      99.4%; Score 171; DB 9; Length 47440;
Best Local Similarity 100.0%; Pred. No. 5.1e-50;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2 GTTCTCAATACCGGAGGAGCAGAGCTATTTCAGCCACATGAAAAAGCATCGGAATTGA 61
DB 19497 GTTCTCAATACCGGAGGAGCAGAGCTATTTCAGCCACATGAAAAAGCATCGGAATTGA 19438
QY      62 GATCGCAGCTCAGAGCAGACCGGGCGCCCTTCACCTTCAGAGAGCTTTGATTTCTTG 121
DB 19437 GATCGCAGCTCAGAGCAGACCGGGCGCCCTTCACCTTCAGAGAGCTTTGATTTCTTG 19378
QY      122 CATCTGGCTGCTGGAGCTTCCTTAGGCGAGTAACAAATACATAAAGCAG 172
DB 19377 CATCTGGCTGCTGGAGCTTCCTTAGGCGAGTAACAAATACATAAAGCAG 19327

```

RESULT 5 *
AL772370/c

```

LOCUS      AL772370      150319 bp      DNA      linear      HTG 17-AUG-2002
DEFINITION Homo sapiens chromosome X clone RP11-164B10, *** SEQUENCING IN
ACCESSION  AL772370
VERSION     AL772370.5 GI:22416024
KEYWORDS   HTG, HTGS_PHASE1, HTGS_ACTIVIFERIN, HTGS_DRAFT, HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgehire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Aug 21, 2002 this sequence version replaced gi:22204612.
            Genome Center
COMMENT     Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            Project: Information
            Center project name: BA184B10
            ----- Summary Statistics -----
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 99% of reads
            Chemistry: Dye-terminator RT-amersham; 0% of reads Consensus
            quality: 149793 bases at least Q40
            Consensus quality: 149914 bases at least Q30
            Consensus quality: 150006 bases at least Q20
            Insert size: 150219; sum-of-contigs
            Insert size: 152259; 2.5% error; agarose-gel
            Quality coverage: 17.18x in Q20 bases; sum-of-contigs Quality
            coverage: 17.58x in Q20 bases; agarose-gel

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 4859: contig of 4859 bp in length
* 4860 4959: gap of 100 bp
* 4960 150319: contig of 145360 bp in length.
Location/Qualifiers
1. 150319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-164B10"
/clone_11b="RP11-11.1"
1. 4859
/note="assembly fragment:05115
fragment chain:1"
4960. 150319
/note="assembly fragment:05270
fragment chain:1
clone_end:SP6
vector side:right"

```

```

FEATURES
misc_feature
misc_feature
misc_feature
BASE COUNT      47436 a 29741 c 29722 g 43320 t 100 others
ORIGIN

```

```

Query Match      99.4%; Score 171; DB 2; Length 150319;
Best Local Similarity 100.0%; Pred. No. 6.7e-50;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2 GTTCTCAATACCGGAGGAGCAGAGCTATTTCAGCCACATGAAAAAGCATCGGAATTGA 61
DB 32269 GTTCTCAATACCGGAGGAGCAGAGCTATTTCAGCCACATGAAAAAGCATCGGAATTGA 32210

```

Oy 62 GATCCGAGCTCAGAGACACCGGGCCCCCTTCCACCTTCCAGAGAGCTTTGATTTCTTG 121
 Db 32209 GATCCGAGCTCAGAGACACCGGGCCCCCTTCCACCTTCCAGAGAGCTTTGATTTCTTG 32150
 Oy 122 CATCTGGCTGCTGGGAGCTTCCCTTAGGACGATTAACAATATATTAAGCAG 172
 Db 32149 CATCTGGCTGCTGGGAGCTTCCCTTAGGACGATTAACAATATATTAAGCAG 32099
 RESULT 6
 AF364071 892 bp mRNA linear ROD 04-MAY-2001
 LOCUS AF364071
 DEFINITION Rattus norvegicus SMPX protein (Smpx) mRNA, complete cds.
 ACCESSION AF364071
 VERSION AF364071.1 GI:13940509
 KEYWORDS
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 892)
 Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.
 Identification, mapping, and genomic structure of a novel
 X-chromosomal human gene (SMPX) encoding a small muscular protein
 Hum. Genet. 105 (5), 506-512 (1999)
 JOURNAL MEDLINE 20065879
 PUBMED 10598820
 REFERENCE 2 (bases 1 to 892)
 Patzak, D.
 Direct Submission
 Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
 Fiechmannstr. 42-44, D-17487 Greifswald, Germany
 JOURNAL
 FEATURES
 source
 1. 892
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 <1. 892
 /gene="Smpx"
 190..447
 /gene="Smpx"
 /codon_start=1
 /product="SMPX protein"
 /protein_id="AAK50399.1"
 /db_xref="GI:13940510"
 /translation="MSKQPSNVSRVIAONINIPMGAPRPGAGPPRRKESTPGTAECA
 PATPEKKVPVPMKKPRGPVNVLSIIONKSELKYVFKGEQ"
 504..509
 /gene="Smpx"
 /evidence=experimental
 862..867
 /gene="Smpx"
 /note="alternate"
 /evidence=experimental
 polyA_signal
 polyA_signal
 862..867
 /gene="Smpx"
 /note="alternate"
 /evidence=experimental
 BASE COUNT 271 a 183 c 193 g 245 t
 ORIGIN
 Query Match 61.9%; Score 106.4; DB 10; Length 892;
 Best Local Similarity 80.1%; Pred. No. 4.6e-27;
 Matches 125; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 Oy 17 AGAGGACAGAGCTATTTCAGCCACATGAAAAGCATCGAATTCGAGCTCAGAG 76
 Db 16 AGAGCACTCCAGCTATTCAGCCACATGAAAAGCATCGAATTCGAGCTCAGAG 75
 Oy 77 GACACCGGGGGCCCCCTTCCACCTTCCAGAGAGCTTTGATTTCTTGATCTGGCTGG 136
 Db 76 GACACCGGGAGTTCTTCTTACCTCTGTAAAGCGTTTTCGTTTTCAGCTGGCTGG 135
 Oy 137 GACTTCCTTAGGACGATTAACAATATATTAAGCAG 172
 Db 136 GACTGTCTCAGGACGATTAACAATATCAGAGAGCAG 171

RESULT 7
 BC005948
 LOCUS BC005948 835 bp mRNA linear PRI 12-JUL-2001
 DEFINITION Homo sapiens, small muscle protein, X-linked, clone MGC:14584
 IMAGE:4246501, mRNA, complete cds.
 ACCESSION BC005948
 VERSION BC005948.1 GI:13543590
 KEYWORDS
 SOURCE MGC.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 835)
 Strauberg, R.
 Direct Submission
 Submitted (02-APR-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA.
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdepax11.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clome distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
 Series: IRAL Plate: 21 Row: a Column: 4
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 6625646.
 FEATURES
 source
 1. 835
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:14584 IMAGE:4246501"
 /issue_type="Skeletal Muscle"
 /clone_lib="NIH MGC_81"
 /lab_host="DH10B"
 /note="Vector: pDNR-LIB"
 113..379
 /codon_start=1
 /product="small muscle protein, X-linked"
 /protein_id="AAH05948.1"
 /db_xref="GI:13543591"
 /translation="NNMSKQPSNVSRVIAONINIPMGAPRPGAGPPRRKESTPGTAECA
 GVPPTSDEKKRTPGAKKLPGPAVNVLSIIONKSELKYVFKAEQ"
 CDS
 113..379
 /codon_start=1
 /product="small muscle protein, X-linked"
 /protein_id="AAH05948.1"
 /db_xref="GI:13543591"
 /translation="NNMSKQPSNVSRVIAONINIPMGAPRPGAGPPRRKESTPGTAECA
 GVPPTSDEKKRTPGAKKLPGPAVNVLSIIONKSELKYVFKAEQ"
 BASE COUNT 282 a 155 c 171 g 227 t
 ORIGIN
 Query Match 58.1%; Score 100; DB 9; Length 835;
 Best Local Similarity 100.0%; Pred. No. 9.2e-25;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy* 73 AGAGGACAGCGGGGGCCCCCTTCCACCTTCCAGAGAGCTTTGATTTCTTGATCTGGCTGC 132
 Db 1 AGAGGACAGCGGGGGCCCCCTTCCACCTTCCAGAGAGCTTTGATTTCTTGATCTGGCTGC 60
 Oy 133 CTGGACTTCCTTAGGACGATTAACAATATATTAAGCAG 172
 Db 61 CTGGACTTCCTTAGGACGATTAACAATATATTAAGCAG 100
 RESULT 8
 AY026524

LOCUS AY026524 787 bp mRNA linear ROD 28-JUN-2001
 DEFINITION Mus musculus muscle-specific protein CSL (Csl) mRNA, complete cds.
 ACCESSION AY026524
 VERSION AY026524.1 GI:14575061
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 787)
 AUTHORS Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C., Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F., Mohun,T. and Harvey,R.P.
 TITLE The small muscle-specific protein Csl modifies cell shape and promotes myocyte fusion in an insulin-like growth factor 1-dependent manner
 JOURNAL J. Cell Biol. 153 (5), 985-998 (2001)
 MEDLINE 21275706
 PUBMED 11381084
 REFERENCE 2 (bases 1 to 787)
 AUTHORS Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C., Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F., Mohun,T. and Harvey,R.P.
 TITLE Direct Submission
 JOURNAL Submitted (01-FEB-2001) Developmental Biology, Victor Chang Cardiac Research Institute, 384 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia
 FEATURES
 source 1..787
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /chromosome="X"
 1..787
 /gene="Csl"
 206..463
 /note="Ortholog of Homo sapiens SMPX"
 /product="muscle-specific protein CSL"
 /protein_id="AAK07682.1"
 /db_xref="GI:14575062"
 /translation="MSKQPSNVRAIQANINIPMGAFRPGAPPRKKESTPETEBSA
 PTTSEKKPIPMKKFPGPVNLSIQNVKSELKPKVEQ"
 BASE COUNT 233 a 167 c 185 g 202 t
 ORIGIN
 Query Match 55.3%; Score 95.2; DB 10; Length 787;
 Best Local Similarity 76.5%; Pred. No. 4,9e-23;
 Matches 130; Conservative 0; Mismatches 38; Indels 2; Gaps 1;
 QY 5 CTCATACCGGAGAGACAGAGCTATTTCAGCCACATGAAGACATCGAATTGAGAT 64
 DB 18 CTGAGAGAGACAGAGACTCCAGCTATTTCAGCCACATGAAGACATCGAATTGAGAT 77
 QY 65 CGCAGCTAGAGAGAACCGGGGGGCCCCCTTCCACTTCCAGAGAGC--TTTGATTCCTGC 122
 DB 78 CCCCGCTAGAGAGACACCGGAGGTCCTTCTATCTCTGTAAGCCGCTTTTGTGTTTGC 137
 QY 123 ATCTGGCTGCTGGAGCTTCCCTTAGAGCAGTAACAATATCATTAAGCAG 172
 DB 138 ACTTGGCGGCTGGAGCTGTCTCTCAGGACGTAAACCAATCCAGAGAGCAG 187

LOCUS AF364070 936 bp mRNA linear ROD 04-MAY-2001
 DEFINITION Mus musculus SMPX protein (Smpx) mRNA, complete cds.
 ACCESSION AF364070
 VERSION AF364070.1 GI:13940507
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 936)
 AUTHORS Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
 TITLE Identification, mapping, and genomic structure of a novel X-chromosomal human gene (SMPX) encoding a small muscular protein
 JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
 MEDLINE 2065879
 PUBMED 10598820
 REFERENCE 2 (bases 1 to 936)
 AUTHORS Patzak,D.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang., Fleischmannstr. 42-44, D-17487 Greifswald, Germany
 FEATURES
 source 1..936
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1..936
 /gene="Smpx"
 199..456
 /gene="Smpx"
 /codon_start=1
 /product="SMPX protein"
 /protein_id="AAK50398.1"
 /db_xref="GI:13940508"
 /translation="MSKQPSNVRAIQANINIPMGAFRPGAPPRKKESTPETEBSA
 PTTSEKKPIPMKKFPGPVNLSIQNVKSELKPKVEQ"
 906..911
 /gene="Smpx"
 /evidence="experimental"
 poly_a_signal
 BASE COUNT 283 a 187 c 203 g 263 t
 ORIGIN
 Query Match 55.3%; Score 95.2; DB 10; Length 936;
 Best Local Similarity 76.5%; Pred. No. 5,1e-23;
 Matches 130; Conservative 0; Mismatches 38; Indels 2; Gaps 1;
 QY 5 CTCATACCGGAGAGACAGAGCTATTTCAGCCACATGAAGACATCGAATTGAGAT 64
 DB 11 CTGAGAGAGACAGAGACTCCAGCTATTTCAGCCACATGAAGACATCGAATTGAGAT 70
 QY 65 CGCAGCTAGAGAGAACCGGGGGGCCCCCTTCCACTTCCAGAGAGC--TTTGATTCCTGC 122
 DB 71 CCCCGCTAGAGAGACACCGGAGGTCCTTCTATCTCTGTAAGCCGCTTTTGTGTTTGC 130
 QY 123 ATCTGGCTGCTGGAGCTTCCCTTAGAGCAGTAACAATATCATTAAGCAG 172
 DB 131 ACTTGGCGGCTGGAGCTGTCTCTCAGGACGTAAACCAATCCAGAGAGCAG 180

RESULT 10
 LOCUS NMU245772 943 bp mRNA linear ROD 12-APR-2001
 DEFINITION Mus musculus mRNA for stretch responsive muscle (X-chromosome)
 ACCESSION NMU245772
 VERSION NMU245772.1 GI:10178962
 KEYWORDS Srmx gene; stretch responsive muscle (X-chromosome).
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 943)
 AUTHORS Kemp,T.J., Sadusky,T.J., Simon,M., Brown,R., Easewood,M., Sassoon,D.A. and Coulton,G.R.
 TITLE Identification of a novel stretch-responsive skeletal muscle gene (srmx)
 JOURNAL Genomics 72 (3), 260-271 (2001)
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2 (bases 1 to 943)
 AUTHORS Kemp,T.J.

TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial College School Of Medicine, 6AF Building, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM

FEATURES
source Location/Qualifiers

1..943
/organism="Mus musculus"
/strain="C57BL/10"
/db_xref="taxon:10090"
/cfeature_type="skeletal muscle"
1..943
/gene="Strmx"
1..214
/gene="Strmx"
215..472
/gene="Strmx"
/codon_start=1
/product="stretch responsive muscle (X-chromosome)"
/protein_id="CAC08493.1"
/db_xref="GI:10178963"
/translation="MSKOPISNVRAIQANINPMGAFRPGAGPPRKESTPETERGA
PTTSEKKRIPGKKRPPVNVLSIQNVKSELKVPKRGQ"
473..943
/gene="Strmx"
535..539
/gene="Strmx"
/function="mRNA deestabilising motif"
585..590
/gene="Strmx"
/function="mRNA deestabilising motif"
658..662
/gene="Strmx"
/function="mRNA deestabilising motif"
706..710
/gene="Strmx"
/function="mRNA deestabilising motif"
740..744
/gene="Strmx"
/function="mRNA deestabilising motif"
779..785
/gene="Strmx"
/function="mRNA deestabilising motif"
909..913
/gene="Strmx"
/function="mRNA deestabilising motif"
923..928
/gene="Strmx"
943
/gene="Strmx"
BAGE COUNT 278 a 210 g 262 t 1 others
ORIGIN
Query Match 55.3%; Score 95.2; DB 10; Length 943;
Best Local Similarity 76.5%; Pred. No. 5.2e-23;
Matches 130; Conservative 0; Mismatches 38; Indels 2; Gaps 1;
QY 5 CTCATATCCGGAGAGGACAGAGCTATTTGAGCCACATGAAAAGCATCGAATTGAGAT 64
DB 27 CTGAGAGAGACACAGACATCTCAGCTATTTGAGCCACATGAAAAGCATCGAATTGAGAT 86
QY 65 CGCAGCTCAGAGGACACCGGCGCCCTTCCACCTTCCAGAGAC--TTGTATCTTGGC 122
DB 87 CCGCGCTCAGAGGACACCGGAGTTCCTTCATCTGTAAGCGCTTTTGTGTTTTGC 146
QY 123 ATCTGGCTGCTGGAGCTTCCCTTAGGAGGATGAACAATACATTAAGCAG 172
DB 147 ACCTGGCGCTGGAGCTGTCTCAGGACATGAACAATCAGAGAGCAG 196

RESULT 11
AL731811/c 239667 bp DNA 1linear HTG 12-AUG-2002
LOCUS AL731811 239667 bp DNA 1linear HTG 12-AUG-2002
DEFINITION Mus musculus chromosome X clone RP23-60A1, *** SEQUENCING IN

ACCESSION AL731811
VERSION AL731811.14 GI:22213676
KEYWORDS HTGS PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 13, 2002 this sequence version replaced gi:21912698.
COMMENT

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: bme0a1

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 96% of reads

Chemistry: Dye-terminator Big Dye; 3% of reads

Consensus quality: 235503 bases at least Q40

Consensus quality: 237530 bases at least Q30

Insert size: 238467; sum-of-contigs

Insert size: 241670; 2.5% error; agarose-fp

Quality coverage: 7.58x in Q20 bases; sum-of-contigs Quality

coverage: 10.48x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 39133: contig of 39133 bp in length
* 39134 39233: gap of 100 bp
* 39234 84228: contig of 44995 bp in length
* 84229 84328: gap of 100 bp
* 84329 106119: contig of 21791 bp in length
* 106120 106219: gap of 100 bp
* 106220 119269: contig of 13050 bp in length
* 119270 119369: gap of 100 bp
* 119370 130432: contig of 11063 bp in length
* 130433 130532: gap of 100 bp
* 130533 218182: contig of 87650 bp in length
* 218183 218282: gap of 100 bp
* 218283 221609: contig of 3327 bp in length
* 221610 221709: gap of 100 bp
* 221710 223956: contig of 2147 bp in length
* 223957 223956: gap of 100 bp
* 223957 226308: contig of 2332 bp in length
* 226309 226408: gap of 100 bp
* 226409 228884: contig of 2476 bp in length
* 228885 228984: gap of 100 bp
* 228985 231639: contig of 2655 bp in length
* 231640 231739: gap of 100 bp
* 231740 236718: contig of 4979 bp in length
* 236719 236819: gap of 101 bp
* 236820 239667: contig of 2848 bp in length.

FEATURES

source

1..239667
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-60A1"

```

misc_feature      /clone.lib="RPCT-23"
                  1..39133
/note="assembly_fragment:09640
fragment_chain:1"
misc_feature      39234..84228
/note="assembly_fragment:09628
fragment_chain:1"
misc_feature      84329..106119
/note="assembly_fragment:09663
fragment_chain:1"
misc_feature      106220..119269
/note="assembly_fragment:04062
fragment_chain:1"
misc_feature      119370..130432
/note="assembly_fragment:01671
fragment_chain:1"
misc_feature      130533..218182
/note="assembly_fragment:09523
fragment_chain:1"
misc_feature      218283..221609
/note="assembly_fragment:00423"
misc_feature      221710..223856
/note="assembly_fragment:05853"
misc_feature      223957..226308
/note="assembly_fragment:07191"
misc_feature      226409..228884
/note="assembly_fragment:08648"
misc_feature      228985..231639
/note="assembly_fragment:08815"
misc_feature      231740..236718
/note="assembly_fragment:09570"
misc_feature      236819..239667
/note="assembly_fragment:09651"
BASE COUNT      72780 a 48417 c 49370 g 67888 t 1212 others
ORIGIN

```

```

Query Match      55.3%; Score 95.2; DB 2; Length 239667;
Best Local Similarity 76.5%; Pred. No. 1.8e-22;
Matches 130; Conservative 0; Mismatches 38; Indels 2; Gaps 1;

QY      5 CTCATACCGGAGAGGACAGAGCTATTTCAGCCACATGAAAGCATCGAATTGAGAT 64
DB 77537 CTGAGAGAGAGACAGACACTCCAGCTATTTCAGCCACATGAAAGCATCGAATTGAGAT 77478

QY      65 CGCAGCTAGAGAGACACCGGGGCCCCCTTCCACCTTCCAGAGAGC--TTGTATTCTTGC 122
DB 77477 CCCCCTCAGAGAGACACCGGAGATTCCTTCTGATCTGTAAGCGCTTTTGTGTTTGC 77418

QY      123 ATCTGCTGCTCGGAGACTTCCCTTAGCAGTAACAATATACATAAGAG 172
DB 77417 ACTTGGCCGCTTGGAGCTGTCTCTCAGGAGTAACCAATCAGAGAGAG 77368

```

```

RESULT 12
AP002022/c      173933 bp      DNA      linear      PRI 01-JUN-2000
LOCUS
DEFINITION      Homo sapiens genomic DNA, chromosome 4q22-q24, clone:2198J12,
complete sequence.
ACCESSION      AP002022
VERSION      AP002022.1 GI:7798582
KEYWORDS
SOURCE
ORGANISM      Homo sapiens DNA, clone:2198J12.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
AUTHORS      Tsai, S.F.
TITLE      Genome Sequencing of the Chromosome 4q Region Implicated in Human
Hepatocellular Carcinoma Pathogenesis 1
JOURNAL
2 (bases 1 to 173933)
REFERENCE
AUTHORS      Tsai, S.F.
TITLE      Direct Submission

```

```

JOURNAL
Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
University, Institute of Genetics, 155 Li-Rong St. Section 2,
Paitou, Taipei, Taiwan 11221, Republic of China
(E-mail: ympei@ym.edu.tw, URL: http://genome.ym.edu.tw/,
Tel: 886-2-28267043, Fax: 886-2-28264930)
COMMENT
Quality: the expected Phred/Phrap calculated error rate (per 10kb)
is 0.01; Estimated total number of errors is 0.03.
FEATURES
Location/Qualifiers
1..173933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q22-q24"
/clone="2198J12"
BASE COUNT      54741 a 34703 c 32692 g 51797 t
ORIGIN

```

```

Query Match      20.1%; Score 34.6; DB 9; Length 173933;
Best Local Similarity 59.8%; Pred. No. 1.3;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY      50 CATCGAATTGAGATCCAGCTCAGAGACACCGGGCCCCCTTCCACCTTCCAGAGAGC 109
DB 77208 CAGTTGAAGAGAGTCACTGGCCCTAAGTCACTCCGACCACTCCACTTCCAAAGAG 77149

QY      110 TTGTATTCTTTCATCTGCTGCTGCGACTTCCCTT 146
DB 77148 ATTCATGTAAAGATATATAAAGCTCAGAGAGTCCCT 77112

```

```

RESULT 13
AC098674
LOCUS
DEFINITION      Homo sapiens chromosome 4 clone RP11-18319, WORKING DRAFT SEQUENCE,
12 unordered pieces.
ACCESSION      AC098674 AC013536
VERSION      AC098674.1 GI:16506917
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM      Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 188818)
AUTHORS      Waterston, R.H.
TITLE      The sequence of Homo sapiens clone
2 (bases 1 to 188818)
JOURNAL
Unpublished
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
COMMENT
Submitted (28-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Oct 28, 2001 this sequence version replaced gi:10047701.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: H_NH0183109
Drafting center: WIBR
----- Summary Statistics -----
Sequencing vector: M13; 40%
Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 181860 bases at least Q40
Consensus quality: 184864 bases at least Q30
Consensus quality: 186595 bases at least Q20
Insert size: 187000; agarose-1p

```


Insert size: 188928; sum-of-Contigs
Quality coverage: 10.12 in Q20 bases; agarose-fp
Quality coverage: 10.23 in Q20 bases; sum-of-Contigs

NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1531: contig of 1531 bp in length
1532 1631: gap of unknown length
1632 2914: contig of 1283 bp in length
2915 3015: gap of unknown length
3015 4291: contig of 1277 bp in length
4292 4392: gap of unknown length
4392 6309: contig of 1917 bp in length
6309 6409: gap of unknown length
6409 7737: contig of 1328 bp in length
7737 7837: gap of unknown length
7837 9590: contig of 1753 bp in length
9590 9690: gap of unknown length
9690 11210: contig of 1520 bp in length
11210 11310: gap of unknown length
11310 12356: contig of 1047 bp in length
12357 12456: gap of unknown length
12457 25733: contig of 13277 bp in length
25734 25833: gap of unknown length
25834 45913: contig of 20080 bp in length
45914 46014: gap of unknown length
46014 78055: contig of 32041 bp in length
78055 78155: gap of unknown length
78155 188818: contig of 110664 bp in length.
Location/Qualifiers

1. 188818
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-18319"
1. 1531
/note="assembly_name:Contig23"
1632. 2914
/note="assembly_name:Contig44"
3015. 4291
/note="assembly_name:Contig48"
4392. 6308
/note="assembly_name:Contig50"
6409. 7736
/note="assembly_name:Contig51"
7837. 9589
/note="assembly_name:Contig52"
9690. 11209
/note="assembly_name:Contig53"
11310. 12356
/note="assembly_name:Contig54"
12457. 25733
/note="assembly_name:Contig56"
25834. 45913
/note="assembly_name:Contig57"
46014. 78054
/note="assembly_name:Contig58
clone_end:17
vector_side:left"
78155. 188818
/note="assembly_name:Contig59
clone_end:596
vector_side:right"
59010 a 37485 c 35625 g 55594 t 1104 others

Query Match 20.1%; Score 34.6; DB 2; Length 188818;

Best Local Similarity 59.8%; Pred. No. 1.3;
Matches 58; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
Qy 50 CATGGAATTGATTCGACGTACGAGACACCGGGCCCTTCCACCTTCCAGAGC 109
Db 28192 CAGTTGAAGAGACGACCTGCGCTACAGTACCTCGACACCTCCCAAGAGG 28251
Qy 110 TTGTGATTTCTTGACATCTGCGTCCCTGAGACTTCCTT 146
Db 28252 ATTCATGTATGAATATTAAGCTGAGAGCTCCCT 28288

RESULT 14
AC110999/c 114594 bp DNA 114594 bp DNA PRI 16-APR-2002
DEFINITION Homo sapiens BAC clone RP13-588120 from 4, complete sequence.
AC110999
VERSION AC110999.4 GI:19807878
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
JOURNAL Sulton, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE
AUTHORS 2 (bases 1 to 114594)
TITLE Abbott, S. and Cotton, M.
The sequence of Homo sapiens BAC clone RP13-588120
JOURNAL Unpublished (2001)
AUTHORS 3 (bases 1 to 114594)
TITLE Waterston, R.H.
JOURNAL Submitted (17-FEB-2002) Genome Sequencing Center, Washington
MO 63108, USA

REFERENCE
AUTHORS 4 (bases 1 to 114594)
TITLE Waterston, R.H.
JOURNAL Submitted (21-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 114594)
TITLE Waterston, R.H.
JOURNAL Submitted (29-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS 6 (bases 1 to 114594)
TITLE Waterston, R.
JOURNAL Submitted (16-APR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 29, 2002 this sequence version replaced gi:19571169.

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
Summary Statistics
Center project name: H_FH0588120

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCR-13 Human Female BAC library was constructed using improved cloning techniques developed by Kazutoyo Osoegawa. The library was generated by Baohui Zhao in our laboratory. Construction was funded by a grant from the National Human Genome Research Institute (NHGRI, NIH) (#R01RG01165-03). The library was generated according to the new NHGRI/DOE 'Guidance on Human Subjects in Large-Scale DNA Sequencing'.

Female blood was obtained via a double-blind selection protocol. Female blood DNA was isolated from one randomly chosen donor (out of 10 female donors) and partially digested with a combination of EcoRI and EcoRI Methylase for library segments 1&2 or either MboI or DpnII for library segments 3&4. Size selected DNA was cloned into the pBACe3.6 vector between the EcoRI sites for library segments 1&2 or the BamHI sites for library segments 3&4. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). The library has been arrayed into 384-well microtiter dishes and also gridded onto 22x22cm nylon high density filters for screening by probe hybridization.

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-621120, 2000 bp overlap. Actual start of this clone is at base position 1 of RP13-588120; actual end is at base position 43117 of RP11-621120.

Data from AC108928 was used to finish this clone, AC110999.

Location/Qualifiers

1. 114594

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP13-588120"

/clone_1lb="RPC1-13"

35. 300

/rpc_family="MIR"

repeat_region

353. 386

/rpc_family="AT-rich"

repeat_region

449. 1050

/rpc_family="L1"

repeat_region

1255. 1461

/rpc_family="Alu"

repeat_region

1255. 1271

/note="similar to Homo sapiens EST A1147140 (NID:93674822)"

misc_feature

OK34606.s1

repeat_region

1598. 1694

/rpc_family="Alu"

repeat_region

2849. 3000

/rpc_family="MIR"

repeat_region

6086. 6108

/rpc_family="(TTG)n"

repeat_region

6295. 6599

/rpc_family="Alu"

repeat_region

repeat_region

9883. 10428

/rpc_family="ERV1"

repeat_region

10466. 10974

/rpc_family="CR1"

repeat_region

11090. 11117

/rpc_family="AT-rich"

repeat_region

12464. 12683

/rpc_family="MIR"

repeat_region

14156. 14315

/rpc_family="MIR"

repeat_region

14665. 14872

/rpc_family="ERV1"

repeat_region

14713. 15018

/rpc_family="ERV1"

repeat_region

15382. 15446

/rpc_family="MIR"

repeat_region

16224. 16273

/rpc_family="(TA)n"

repeat_region

20970. 21006

/rpc_family="(GTTG)n"

repeat_region

23556. 24442

/rpc_family="L2"

repeat_region

25249. 25420

/rpc_family="MIR"

repeat_region

25460. 25601

/rpc_family="Alu"

repeat_region

25908. 25989

/rpc_family="L1"

repeat_region

26579. 26683

/rpc_family="L2"

repeat_region

27799. 27958

/note="match to EST AA354508 (NID:92007063)"

misc_feature

27851. 27958

/note="similar to Mus musculus EST BB646518 (NID:916480847)"

repeat_region

30265. 30643

/rpc_family="MALR"

repeat_region

31684. 31732

/rpc_family="AT-rich"

repeat_region

32156. 32197

/rpc_family="(TAAA)n"

repeat_region

33275. 33347

/note="similar to Mus musculus EST BB646518 (NID:916480847)"

misc_feature

33275. 33347

/note="match to EST AA354508 (NID:92007063)"

repeat_region

33754. 33781

/rpc_family="AT-rich"

repeat_region

33906. 34198

/rpc_family="Alu"

repeat_region

35243. 35271

repeat_region

35668. 35777

/rpc_family="MER1_type"

repeat_region

36724. 36752

/rpc_family="AT-rich"

repeat_region

37736. 38024

/rpc_family="Alu"

repeat_region

39982. 41119

/rpc_family="L1"

repeat_region

41130. 41171

/rpc_family="A-rich"

repeat_region

41236. 41819

/note="match to EST BB217886 (NID:98905204) hv31b07.x1"

repeat_region

41525. 41747

/rpc_family="L2"

repeat_region

42174. 42261

/rpc_family="MER2_type"

repeat_region

19.5%; Score 33.6; DB 9; Length 114594;

Best Local Similarity 59.4%; Pred. No. 2.7;

Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

41 CATGAAAGCATGCAATTCAGATCGAGTCAAGACACCGGCGCCCTTCCACTT 100
 Db 110981 CATGAAAGCATGCAATTCAGATCGAGTCAAGACACCGGCGCCCTTCCACTT 110922
 101 CCAAGAGCTTTGTTCTTCTGATCGCTGCTGCTG 136
 Db 110921 CCAAGAGCTTTGTTCTTCTGATCGCTGCTGCTG 110886

RESULT 15
 AC079954 158785 bp DNA linear PRI 25-JUL-2002
 LOCUS AC079954
 DEFINITION Human sapiens 12q BAC RP11-179A1 (Roswell Park Cancer Institute
 accession AC079954 complete sequence.
 VERSION AC079954
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 158785)

REFERENCE
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Al-Osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amaral, H.C., Are, J.R., Banks, T., Barbata, J.,
 Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouch, J.,
 Bowls, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
 Chen, Z., Chavez, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Cox, C., Chiu, D., Chowdhury, I., Christopoulos, C., Cleveland, C.D.,
 Davy, Carroll, L., Dedrich, D.A., Delaney, K.R., Delgado, O.,
 Dem, A.L., Ding, Y., Dink, H.H., Douthett, K.J., Draper, H.,
 Dugan-Roches, S., Durbin, K.J., Earmhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Emelting, S., Escotto, M., Falis, T., Ferraguto, D.,
 Flagg, N., Ford, J., Foster, P., Frantz, P., Gabriel, A., Gao, J.,
 Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guerra, W.,
 Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K.,
 Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O.,
 Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F.,
 Howard, S., Huber, J., Hulys, S., Hume, J., Ioshikhes, I., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, J., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, L.,
 Kratoch, J., Kureishi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtenberg, O., Lieu, C., Liu, J., Liu, W.,
 Louis, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapa, P., Maron, I., Martin, R.,
 Martindale, A., Martinez, E., Massey, B., Mawhinney, E., McLeod, M.P.,
 Meador, M., Mei, G., Metzger, S., Metzger, M., Miller, A., Miner, G.,
 Moritz, S., Moser, M., Neal, D., Nelson, D., Newton, J., Morgan, M.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S.,
 Ogihara, M., Okunishi, G., Ogunyemi, N., Oviado, R., Paine, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Rutz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C.,
 Shoenberger, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A.,
 Stanley, H., Stone, H., Sutton, A., Swalek, A., Taber, P., Tamerisa, A.,
 Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B.,
 Vinson, R., Wall, R., Wang, S., Ward, Moore, S., Warren, V., Villalón, D.,
 Washington, C., Watlington, S., Williams, G., Williamson, A.,
 Wleciak, R., Woodard, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J.,
 Zorilla, S., Kuchelapatti, R. and Gibbs, R.
 Direct Submission
 2 (bases 1 to 158785)
 Submitted (20-SEP-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 158785)
 Worley, K.C.

TITLE
 JOURNAL
 Submitted (01-MAY-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 158785)
 Worley, K.C.
REFERENCE
 JOURNAL
 Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 5 (bases 1 to 158785)
 Worley, K.C.
REFERENCE
 AUTHORS
 JOURNAL
 Submitted (25-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 1, 2001 this sequence version replaced g1:13877202.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-4) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Summary Statistics
 Contig length: 158785
 Phrap values in estimate: 158572
 Average error rate (BCM-Phrap estimate): 4.46022e-06
 Fraction of Phrap values less than 40 : 0.00425044
 Number of consensus changing edits: 30
 Number of N's in consensus : 0

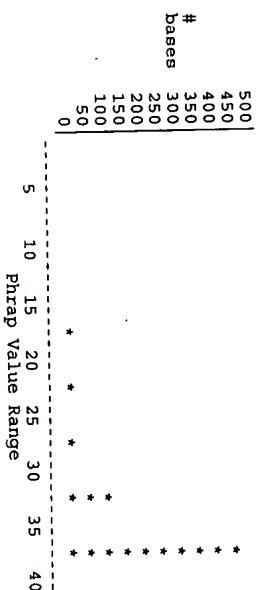
Consensus changing edits
 Position Original+Context Edited+Context
 10464 cctctctc(n)tcnctctc cctctctc(c)tcctctctc
 10467 cctctctc(n)tcnctctc cctctctc(c)tcctctctc
 10469 cctctctc(n)tcctctctc cctctctc(c)tcctctctc
 10491 cctctctc(n)tcctctctc cctctctc(c)tcctctctc
 10511 cctctctc(n)tcctctctc cctctctc(c)tcctctctc

```

13930      taccaccac (n) nnnnnncaag      taccaccac (c) ctgcacaag
13931      accaccacn (n) tnnnncaaga      accaccac (c) tgcacaaga
13932      ccaccacnt (n) nnnncaagaac      ccaccac (c) caacaagaac
13933      ccaactcmn (n) nnnncaagaac      ccaactcc (g) caacaagaac
13934      cactcmn (n) nnnncaagaac      cactcc (c) caacaagaac
13935      actcmn (n) nnnncaagaac      actcc (a) caacaagaac
13936      aactaaalt (n) ggaantagc      aactaaalt (t) ggaantagc
13937      attnggcaa (n) tagatccaa      attnggcaa (g) tagatccaa
13938      aacaatat (n) ggaatccaa      aacaatat (t) ggaatccaa
13939      tatgactaa (n) ggaatccaa      tatgactaa (t) ggaatccaa
13940      gacccaatag (n) ggaatccaa      gacccaatag (t) ggaatccaa
13941      aagtcctc (n) aagaagtc      aagtcctc (g) aagaagtc
13942      actccatc (n) aagaagtc      actccatc (t) aagaagtc
13943      gtaattcag (n) tctttttt      gtaattcag (t) tctttttt
13944      ctagnctt (n) tctttttt      ctagnctt (g) tctttttt
13945      actgaccca (n) ccaatccaa      actgaccca (t) ccaatccaa
13946      ctatcttg (n) aacctgtaa      ctatcttg (g) aacctgtaa
13947      aagtgggga (n) ggcctttaa      aagtgggga (t) ggcctttaa
13948      ctgttga (n) ggcctttaa      ctgttga (g) ggcctttaa
13949      acccaatgt (n) tggatcagc      acccaatgt (t) tggatcagc
13950      aacgcaaga (n) naacaagtc      aacgcaaga (g) naacaagtc
13951      acagcaaga (n) naacaagtc      acagcaaga (t) naacaagtc
13952      tggatcag (n) ggcacatc      tggatcag (g) ggcacatc
13953      gtaattag (n) agagacgag      gtaattag (t) agagacgag
13954      ctccagaga (a) t. . . . . ctccagaga (t) t. . . . .

```

----- Distribution of Quality < 40 Bases -----



Version: 1.01 gxf.

Location/Qualifiers

Source

```

1. 158785
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="12"
   /clone="RP11-179A1"
1. 1977
   /note="overlaps bases 105124..107100 of clone AC117377"
   /function="clone overlap"
1375..1611
   /note="Region similar to Hs#S1466340 w/f36n03.x1 Homo
sapiens cDNA: A1806918"
1612..2413
   /rpt_family="L2"
3529..3665
   /rpt_family="FLAM_C"
3892..4025
   /rpt_family="MIR"

```

Query Match

19.1%: Score 32.8; DB 9; Length 158785;

Best Local Similarity 64.5%: Pred. No. 5.8; Mismatches 27; Indels 0; Gaps 0;

Matches 49; Conservative 0;

```

Oy      80 ACCGGGCGCCCTTC 95
Db      25443 ACAGAGACTTCTTAC 25458

Search completed: April 15, 2003, 20:53:11
Job time : 618.696 secs

```

```

Oy      20 GGACAGAGCTATTTCAGCAGCAGTGAAGATTCGATTCGAGCTCAGAGAGAC 79
Db      25383 GCCTGTAGCAGCTAAGACAGATTAAGACAGACAAAGGAGCTCGGTGAC 25442

```

```

RESULT 7
AA211521
LOCUS
DEFINITION zns5b01.r1 Stragatene muscle 937209 Homo sapiens cDNA clone
ACCESSION IMAGE:562057 5', mRNA sequence.
VERSION AA211521
KEYWORDS AA211521.1 GI:1810175
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 501)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9):807-828 (1996)
9704478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 443.
FEATURES
Location/Qualifiers
1..501
/organism="Homo sapiens"
/db_xref="CDB:4595347"
/db_xref="taxon:96069"
/clone="IMAGE:562057"
/clone_lib="Stragatene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site 1: ECORI; Site 2: XhoI; Cloned unidirectionally.
Primer: Oligo dt. Skeletal muscle from patient with
malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 167 a 103 c 116 g 115 t
ORIGIN
Query Match 100.0%; Score 57; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
|||||
Db 86 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 142

RESULT 8
BM697544
LOCUS
DEFINITION UI-E-DXO-agn-i-12-0-UI.r1 UI-E-DXO Homo sapiens cDNA clone
ACCESSION UI-E-DXO-agn-i-12-0-UI 5', mRNA sequence.
VERSION BM697544
KEYWORDS BM697544.1 GI:19010802
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 719)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9):807-828 (1996)
9704478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 443.
FEATURES
Location/Qualifiers
1..719
/organism="Homo sapiens"
/db_xref="CDB:4595347"
/db_xref="taxon:96069"
/clone="IMAGE:562057"
/clone_lib="Stragatene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site 1: ECORI; Site 2: XhoI; Cloned unidirectionally.
Primer: Oligo dt. Skeletal muscle from patient with
malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 167 a 103 c 116 g 115 t
ORIGIN
Query Match 100.0%; Score 57; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
|||||
Db 86 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 142

RESULT 9
BF791178
LOCUS
DEFINITION 602251278F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4338594 5',
mRNA sequence.
ACCESSION BF791178
VERSION BF791178.1 GI:12096232
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 756)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

```

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 DNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES

source

1..719
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-DXO-agn-i-12-0-UI"
 /clone_lib="UI-E-DXO"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DXO is a cDNA library containing the following
 tissues(s): fetal eyes. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into p773-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is AGAATCAGA. This library
 was created for the program Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."
 BASE COUNT 226 a 142 c 163 g 186 t
 ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

100.0%; Score 57; DB 14; Length 719;
 100.0%; Pred. No. 4.2e-10;
 57; 0; 0; 0; 0;

QY

Db

1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
 |||||
 128 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 184

RESULT 9

BF791178

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BF791178
 602251278F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4338594 5',
 mRNA sequence.
 BF791178
 BF791178.1 GI:12096232
 EST.
 human.
 Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 756)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 18:42:01 ; Search time 3.37265 Seconds
(without alignments)
5183.040 Million cell updates/sec

Title: US-09-647-019-7
Perfect score: 57
Sequence: 1 ggataagactgcgaatat.....ccaaatgtagccatccag 57

Scoring table: IDENTITY NUC
Gap 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.4	77.9	909	4	US-09-484-970B-111
2	23.2	40.7	1539	4	US-09-668-680-13
3	21.8	38.2	519	3	US-09-284-782-22
4	21.8	38.2	246240	2	US-08-724-394A-20
5	21.8	38.2	246240	2	US-08-724-394A-21
6	21.8	38.2	246240	2	US-08-724-394A-22
7	21.2	37.2	5093	1	US-08-468-036-23
8	21.2	37.2	5093	2	US-08-376-843-23
9	21.2	37.2	246240	2	US-08-724-394A-20
10	21.2	37.2	246240	2	US-08-724-394A-21
11	21.2	37.2	246240	2	US-08-724-394A-22
12	20.8	36.5	1357	1	US-08-370-975B-8
13	20.8	36.5	5113	3	US-08-973-273-2
14	20.8	36.5	26764	1	US-08-370-975B-1
15	20.6	36.1	361	1	US-08-094-079-12
16	20.6	36.1	361	1	US-08-094-079-12
17	20.6	36.1	361	1	US-08-094-079-18
18	20.6	36.1	361	1	US-08-094-079-20
19	20.6	36.1	361	1	US-08-094-079-22
20	20.6	36.1	514	1	US-08-094-079-6
21	20.6	36.1	514	1	US-08-094-079-10
22	20.6	36.1	630	5	PCT-US92-06617A-10
23	20.6	36.1	5125	1	US-08-094-948A-4
24	20.6	36.1	5125	5	PCT-US96-09319-4
25	20.4	35.8	4011	1	US-08-121-057-3
26	20.4	35.8	4011	2	US-08-509-187D-3
27	20.4	35.8	4011	2	US-09-121-396-3

C 28	20.4	35.8	4011	5	PCT-US93-09704A-3	Sequence 3, Appli
C 29	20.4	35.8	4079	1	US-08-121-057-2	Sequence 2, Appli
C 30	20.4	35.8	4079	2	US-08-509-187D-2	Sequence 2, Appli
C 31	20.4	35.8	4079	2	US-09-121-396-2	Sequence 2, Appli
C 32	20.4	35.8	4079	5	PCT-US93-09704A-2	Sequence 2, Appli
C 33	20.4	35.8	5854	4	US-09-221-017B-497	Sequence 497, App
C 34	20.2	35.4	1521	1	US-08-670-354-3	Sequence 3, Appli
C 35	20.2	35.4	1521	4	US-09-320-424-3	Sequence 3, Appli
C 36	20.2	35.4	1521	5	PCT-US96-10895-3	Sequence 3, Appli
C 37	20.2	35.4	2436	4	US-08-983-275-1	Sequence 1, Appli
C 38	20	35.1	2007	4	US-09-134-001C-593	Sequence 593, App
C 39	19.8	34.7	299	4	US-09-328-111-349	Sequence 349, App
C 40	19.8	34.7	427	1	US-08-334-254-10	Sequence 10, Appl
C 41	19.8	34.7	427	2	US-08-848-131-10	Sequence 10, Appl
C 42	19.8	34.7	427	5	PCT-US95-14792-10	Sequence 10, Appl
C 43	19.8	34.7	678	4	US-09-221-017B-345	Sequence 345, App
C 44	19.8	34.7	986	1	US-07-637-250A-8	Sequence 8, Appli
C 45	19.8	34.7	986	1	US-08-145-061-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-484-970B-111
; Sequence 111, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne G.
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 111
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 021656.2CBI
US-09-484-970B-111

Query Match 77.9%; Score 44.4; DB 4; Length 909;
Best Local Similarity 96.6%; Pred. No. 1.3e-08;
Matches 56; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 GGATAAGACTGCATG-ATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
DB 209 GGATAAGACTGCATGAAAAAATGCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 266

RESULT 2
US-09-668-680-13/c
; Sequence 13, Application US/09668680
; Patent No. 6436703
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Xu, Chongjun
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. 6436703el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 790CIP2A
; CURRENT APPLICATION NUMBER: US/09/668,680

1 CURRENT FILING DATE: 2000-09-22
2 PRIOR APPLICATION NUMBER: 09/649,167
3 PRIOR FILING DATE: 2000-08-23
4 PRIOR APPLICATION NUMBER: 09/540,217
5 PRIOR FILING DATE: 2000-03-31
6 NUMBER OF SEQ ID NOS: 13
7 SOFTWARE: pt_FL_genes Version 2.0
8 SEQ ID NO 13
9 LENGTH: 1539
10 TYPE: DNA
11 ORGANISM: Homo sapiens
12 FEATURE:
13 NAME/KEY: CDS
14 LOCATION: (130)..(1539)
15 US-09-668-680-13

Query Match 40.7%; Score 23.2; DB 4; Length 1539;
Best Local Similarity 65.4%; Pred. No. 3.5;
Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1 GGATAAGACTGCATGAATATGCGAAACAGCCAGTTTCCATGTTAGGCCA 52
DB 1365 GGAGAAGAGCAGGAGTGTAAGAAACAGGATGATGCGCATGTTGGGAGCCA 1314

RESULT 3
US-09-284-782-22/c
Sequence 22, Application US/09284782
Patent No. 6057111
GENERAL INFORMATION:
APPLICANT: ENTERPRISES, LTD., QBI
APPLICANT: Deiss, Louis P.
APPLICANT: Yehiely, Fruma
APPLICANT: Efimova, Elena
APPLICANT: Vasquez-Iaslop, No. 6057111a C.
APPLICANT: Binat, Paz
TITLE OF INVENTION: GENE IDENTIFICATION METHOD
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6057111thwestern Highway, Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/284,782.
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Montgomery, Ilene N.
REGISTRATION NUMBER: 38,972
REFERENCE/DOCKET NUMBER: 0168-00022
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 519 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-284-782-22

Query Match 38.2%; Score 21.8; DB 3; Length 519;
Best Local Similarity 78.8%; Pred. No. 9.6;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 19 ATGTGGAACAGCCAGTTTCCATGTTAGGCC 51
DB 392 ATTTGTATCAGCCAGTTTCCCTTGTAGGCC 360

RESULT 4
US-08-724-394A-20/C
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-20
Query Match 38.2%; Score 21.8; DB 2; Length 246240;
Best Local Similarity 78.8%; Pred. No. 42;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 GATAAGACTGCATGAATATGTCGAAACAGCCAG 34
DB 65491 GATCAGACTTCAGGAATATGTTGACAAAGGCAG 65459

RESULT 5
US-08-724-394A-21/c
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.


```

: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitts, Renee A.
: REGISTRATION NUMBER: 35,136
: REFERENCE/DOCKET NUMBER: 017957-000100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-576-0200
: TELEFAX: 415-576-0300
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 246240 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 1..246240
: OTHER INFORMATION: /note= "HLA-H.CONTIG"
: US-08-724-394A-22

Query Match 38.2%; Score 21.8; DB 2;
Best Local Similarity 78.8%; Pred. No. 42;
Matches 26; Conservative 0; Mismatches 7;

Qy 2 GATAGAGCTGCATGATATGTCGAACAGCCAG 34
    ||| ||||| ||||| ||||| ||||| |||||
Db 65491 GATGAGACTTCAGGAATATGTTGACAAAGCCAG 65459

RESULT 7
US-08-468-036-23
: Sequence 23, Application US/08468036
: Patent No. 5728806
: GENERAL INFORMATION:
: APPLICANT: DeMaggio, Anthony J.
: APPLICANT: Hoekstra, Marl P.
: TITLE OF INVENTION: Materials and Methods Relating
: TO THE INVENTION: Interact with Casein Kinase
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSER: Marshall, O'Toole, Gerstein, Murray
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/468,036
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/184,605
: FILING DATE: 21-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5728806and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 278666/31784

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 5093 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-036-23

Query Match 37.2%; Score 21.2; DB 1; Length 5093;
Best Local Similarity 64.0%; Pred. No. 29;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 3 ATAAGACTGCATGAATATGCGAAACAGCCAGTTCCCAATGTTAGAGCCA 52
DB 3902 AAAAACTTCATCAATAGGGGATATAGCCATTGTAATAATTTGTATCA 3951

RESULT 8
US-08-376-843-23
Sequence 23, Application US/08376843
Patent No. 5846764
GENERAL INFORMATION:
APPLICANT: DeMaggio, Anthony J.
APPLICANT: Hoekstra, Merl F.
TITLE OF INVENTION: Materials and Methods Relating to Proteins
TITLE OF INVENTION: that Interact with Casein Kinase I
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/376,843
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,605
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: No. 5846764and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31784
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 5093 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-376-843-23

Query Match 37.2%; Score 21.2; DB 2; Length 5093;
Best Local Similarity 64.0%; Pred. No. 29;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 3 ATAAGACTGCATGAATATGCGAAACAGCCAGTTTCCCAATGTTAGAGCCA 52

DB 3902 AAAAACTTCATCAATAGGGGATATAGCCATTGTAATAATTTGTATCA 3951
RESULT 9
US-08-724-394A-20
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20
Query Match 37.2%; Score 21.2; DB 2; Length 246240;
Best Local Similarity 69.0%; Pred. No. 72;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 4 TAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCCAATGTT 45
DB 222283 TAAGTCCTTAGGAATATGCTCGAAGAGCTAATTTTCATATGTT 222324
RESULT 10
US-08-724-394A-21
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.

```

; APPLICANT: Thomas, Winston
; APPLICANT: Teuchinashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
;
; US-08-724-394A-21
;
; Query Match 37.2%; Score 21.2; DB 2; Length 246240;
; Best Local Similarity 69.0%; Pred. No. 72;
; Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
;
; QY 4 TAAGACTGCATGATATGTCGAACACGCCAGTTTCCCAATGTT 45
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 222283 TAAGCTCTTAGGAATATGCTAGAGAGCTAATTTTCATATGTT 222324
;
; RESULT 11
; US-08-724-394A-22
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Teuchinashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSES: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
;
; Query Match 37.2%; Score 21.2; DB 2; Length 246240;
; Best Local Similarity 69.0%; Pred. No. 72;
; Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
;
; QY 4 TAAGACTGCATGATATGTCGAACACGCCAGTTTCCCAATGTT 45
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 222283 TAAGCTCTTAGGAATATGCTAGAGAGCTAATTTTCATATGTT 222324
;
; RESULT 12
; US-08-370-975B-8/c
; Sequence 8, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H. CONTIG"
;
; US-08-724-394A-22
;
; Query Match 37.2%; Score 21.2; DB 2; Length 246240;
; Best Local Similarity 69.0%; Pred. No. 72;
; Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
;
; QY 4 TAAGACTGCATGATATGTCGAACACGCCAGTTTCCCAATGTT 45
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; Db 222283 TAAGCTCTTAGGAATATGCTAGAGAGCTAATTTTCATATGTT 222324
;
; RESULT 12
; US-08-370-975B-8/c
; Sequence 8, Application US/08370975B
; Patent No. 5622851
; GENERAL INFORMATION:
; APPLICANT: Maley, Frank
; APPLICANT: Maley, Gladys F.
; APPLICANT: Weiner, Karen X.B.
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,975B
; FILING DATE: 10-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20894/80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716)263-1636
; TELEFAX: (716)263-1600
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

```

Matches	25;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps
QY	16	AATATGTGCAAAACAGCCAGTTTCCAATGTTAG	47					
Db	4571	AGTTGGTCAAAATAGCCAGTTTCTCATGTGCG	4602					
RESULT 14								
US-08-370-975B-1/c								
; Sequence 1, Application US/08370975B								
; Patent No. 5622851								
; GENERAL INFORMATION:								
; APPLICANT: Maley, Frank								
; APPLICANT: Malety, Gladys F.								
; APPLICANT: Weiner, Karen X.B.								
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene								
; NUMBER OF SEQUENCES: 14								
; CORRESPONDENCE ADDRESS:								
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle								
; STREET: Clinton Square, P.O. Box 1051								
; CITY: Rochester								
; STATE: New York								
; COUNTRY: USA								
; ZIP: 14603								
; COMPUTER READABLE FORM:								
; MEDIUM TYPE: Floppy disk								
; COMPUTER: IBM PC compatible								
; OPERATING SYSTEM: PC-DOS/MS-DOS								
; SOFTWARE: PatentIn Release #1.0, Version #1.25								
; CURRENT APPLICATION DATA:								
; APPLICATION NUMBER: US/08/370,975B								
; FILING DATE: 10-JAN-1995								
; CLASSIFICATION: 435								
; ATTORNEY/AGENT INFORMATION:								
; NAME: Timian, Susan J.								
; REGISTRATION NUMBER: 34,103								
; REFERENCE/DOCKET NUMBER: 20894/80								
; TELECOMMUNICATION INFORMATION:								
; TELEPHONE: (716)263-1636								
; TELEFAX: (716)263-1600								
; INFORMATION FOR SEQ ID NO: 1:								
; SEQUENCE CHARACTERISTICS:								
; LENGTH: 26764 base pairs								
; TYPE: nucleic acid								
; STRANDEDNESS: single								
; TOPOLOGY: linear								
; MOLECULE TYPE: DNA (genomic)								
; POSITION IN GENOME:								
; CHROMOSOME/SEGMENT: 4q35								
US-08-370-975B-1								
Query Match	36.5%;	Score 20.8;	DB 1;	Length 26764;				
Best Local Similarity	64.6%;	Pred. No. 62;						
Matches	31;	Conservative	0;	Mismatches	17;	Indels	0;	Gaps
QY	2	GATAAGACTGCATGAATATGTGCAAAACAGCCAGTTTCCAATGTTAG	49					
Db	22926	GATGTGACAGTTTGAAAAAGGCCAATAATTTCAGAATACAATGGTGTAG	22879					
RESULT 15								
US-08-094-079-12/c								
; Sequence 12, Application US/08094079								
; Patent No. 5512545								
; GENERAL INFORMATION:								
; APPLICANT: COOK, Anne L								
; APPLICANT: CRAIG, Stewart								
; APPLICANT: CLEMENTS, John M								
; APPLICANT: EDWARDS, Richard M								
; APPLICANT: BROWN, David								
; TITLE OF INVENTION: PDGF-B ANALOGUES								
; NUMBER OF SEQUENCES: 22								
; CORRESPONDENCE ADDRESS:								

ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 S. Wacker Dr.
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/094,079
FILING DATE: 24-JAN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/GB92/00141
FILING DATE: 24-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9101645.1
FILING DATE: 24-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, John J.
REGISTRATION NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 93,640
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 12:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..361
OTHER INFORMATION: /note= "PDGF-B gene adapted at 5'
OTHER INFORMATION: end for fusion to alpha factor and truncated at 3'
OTHER INFORMATION: end to remove C-terminal coding residues"
US-08-094-079-12

Query Match 36.1%; Score 20.6; DB 1; Length 361;
Best Local Similarity 67.4%; Pred. No. 26;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 12 CATGAATATGTCGAACAGCCAGTTTCATGTTAGAGCCATC 54
Db 209 CCTGAGTAGGTCGACACTGAACGTTTCTGTTGTACAGCAACC 167

Search completed: April 15, 2003, 22:52:13
Job time : 328.373 secs

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 19:13:41 ; Search time 5.69553 Seconds
(without alignments)
8778.558 Million cell updates/sec

Title: US-09-647-019-7
Perfect score: 57
Sequence: 1 ggataagactgcgatgaatat.....ccaatgttagagccatccag 57

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 43858390 residues
Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :		Published Applications NA:*	
1:	/cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*		
2:	/cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*		
3:	/cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*		
4:	/cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*		
5:	/cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*		
6:	/cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*		
7:	/cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*		
8:	/cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*		
9:	/cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*		
10:	/cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*		
11:	/cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*		
12:	/cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*		
13:	/cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*		
14:	/cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	100.0	824	10	US-09-880-192-4
2	48	84.2	466	10	US-09-960-352-5216
3	24	42.1	31314	10	US-09-764-877-3875
C 4	23.2	40.7	533	9	US-09-764-872-715
C 5	23.2	40.7	957	10	US-09-886-055-102
C 6	22.8	40.0	7043	10	US-09-070-927A-288
7	22.6	39.6	2130	10	US-09-917-800A-1601
8	22.6	39.6	2173	9	US-09-898-234-14
9	22.6	39.6	2173	9	US-09-899-429A-24
10	22.6	39.6	2173	9	US-09-792-356-14
11	22.6	39.6	2173	10	US-09-899-422-14
C 12	22	38.6	4056	9	US-09-784-554B-3
C 13	21.8	38.2	653	10	US-09-833-381-1348
C 14	21.6	37.9	373	10	US-09-974-300-3301
15	21.6	37.9	559	10	US-09-864-761-13274
16	21.6	37.9	2536	10	US-09-817-310-1
17	21.6	37.9	397658	10	US-09-813-320-3
18	21.4	37.5	402	10	US-09-817-607-41
19	21	36.8	404	10	US-09-969-347-31

ALIGNMENTS

RESULT 1
US-09-880-192-4
; Sequence 4, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkman, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 3601719CB1
US-09-880-192-4

Query Match 100.0%; Score 57; DB 10; Length 824;
Best Local Similarity 100.0%; Pred. No. 5.6e-13;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 57
DB 206 GGATAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 262

RESULT 2

US-09-960-352-5216
; Sequence 5216, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 5216

; LENGTH: 466

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 23-BOVMS1-014-Q1-E1-E3

US-09-960-352-5216

Query Match

Best Local Similarity 84.28; Score 48; DB 10; Length 466;

Matches 51; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCATCCA 56

Db 91 GGATAAGACCGCTCAGTATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCATCCA 146

RESULT 3

US-09-764-877-3875

; Sequence 3875, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3875

; LENGTH: 31314

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-877-3875

Query Match

Best Local Similarity 42.18; Score 24; DB 10; Length 31314;

Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 5 AAGACTGCATGAATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCA 52

Db 1967 AAATCTGCAAAAATTTGGTGAAAAAGAGAGTCTGCAGTGTTTAAGCCA 2014

RESULT 4

US-09-764-872-715/c

; Sequence 715, Application US/09764872

; Publication No. US20030050231A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PA125

; CURRENT APPLICATION NUMBER: US/09/764,872

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 957

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 715

; LENGTH: 533

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-872-715

Query Match

Best Local Similarity 40.78; Score 23.2; DB 9; Length 533;

Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 6 AGACTGCATGAATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57

Db 532 AGAATATATGAAAAGGTGTACAGAACTTCCAAAACCTAAAAGGCATCCAG 481

RESULT 5

US-09-886-055-102/c

; Sequence 102, Application US/09886055

; Patent No. US20020132273A1

; GENERAL INFORMATION:

; APPLICANT: STRYER, LUBERT

; APPLICANT: ZOZULYA, SERGEY

; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND

; TITLE OF INVENTION: BIOSENSORS OF CHEMICAL SENSANTS

; FILE REFERENCE: 078003-0277150

; CURRENT APPLICATION NUMBER: US/09/886,055

; CURRENT FILING DATE: 2001-06-22

; PRIOR APPLICATION NUMBER: 60/213,812

; PRIOR FILING DATE: 2000-06-22

; NUMBER OF SEQ ID NOS: 522

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 102

; LENGTH: 957

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-886-055-102

Query Match

Best Local Similarity 40.78; Score 23.2; DB 10; Length 957;

Matches 34; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCA 52

Db 783 GGAGAAGAGAGCGGATGTAGAAAACAGGATGATGCCAATGTGGAGCCA 732

RESULT 6

US-09-070-927A-288/c

; Sequence 288, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch

; APPLICANT: Steven Barash

; APPLICANT: Patrick J. Dillon

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/046,655

; FILING DATE: 1997-05-16

; APPLICATION NUMBER: 60/044,031

; FILING DATE: 1997-05-06

; APPLICATION NUMBER: 60/066,009

; FILING DATE: 1997-11-14

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PB369


```
/
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 288:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 7043 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 288:
US-09-070-927A-288

Query Match 40.0%; Score 22.6; DB 10; Length 7043;
Best Local Similarity 66.0%; Pred. No. 41;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 7 GACTGCATGAATATGTCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCA 56
Db 2081 GATGACATGAATGTTTAAACACAGTACTCTTTCTAAATAATTAAGTCATTCA 2032

RESULT 7
US-09-917-800A-1601
/ Sequence 1601, Application US/09917800A
/ Patent No. US20020119462A1
/ GENERAL INFORMATION:
/ APPLICANT: Mendrick, Donna
/ APPLICANT: Porter, Mark
/ APPLICANT: Johnson, Kory
/ APPLICANT: Castle, Arthur
/ APPLICANT: Elashoff, Michael
/ APPLICANT: Gene Logic, Inc.
/ TITLE OF INVENTION: Molecular Toxicology Modeling
/ FILE REFERENCE: 44921-5038-US
/ CURRENT APPLICATION NUMBER: US/09/917,800A
/ CURRENT FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: US 60/222,040
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: US 60/222,880
/ PRIOR FILING DATE: 2000-11-02
/ PRIOR APPLICATION NUMBER: US 60/290,029
/ PRIOR FILING DATE: 2001-05-11
/ PRIOR APPLICATION NUMBER: US 60/290,645
/ PRIOR FILING DATE: 2001-05-15
/ PRIOR APPLICATION NUMBER: US 60/292,336
/ PRIOR FILING DATE: 2001-05-22
/ PRIOR APPLICATION NUMBER: US 60/295,798
/ PRIOR FILING DATE: 2001-06-06
/ PRIOR APPLICATION NUMBER: US 60/297,457
/ PRIOR FILING DATE: 2001-06-13
/ PRIOR APPLICATION NUMBER: US 60/298,884
/ PRIOR FILING DATE: 2001-06-19
/ PRIOR APPLICATION NUMBER: US 60/303,459
/ PRIOR FILING DATE: 2001-07-09
/ NUMBER OF SEQ ID NOS: 1740
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1601
/ LENGTH: 2130
/ TYPE: DNA
/ ORGANISM: Rattus norvegicus
/ FEATURE:
/ OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_013091
US-09-917-800A-1601

Query Match 39.6%; Score 22.6; DB 10; Length 2130;
Best Local Similarity 64.2%; Pred. No. 32;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 5 AAGACTGCATGAATATGTCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
Db 803 AGGAATGTAATGAAGCTGTGCTTACCTCCAGTTGCAATGTCACAAACCCCCAG 855

RESULT 8
US-09-898-234-14
/ Sequence 14, Application US/09898234
/ Patent No. US20020155112A1
/ GENERAL INFORMATION:
/ APPLICANT: Hauptmann, Rudolph
/ APPLICANT: Himmler, Adolph
/ APPLICANT: Maurer-Fogy, Ingrid
/ APPLICANT: Stratowa, Christian
/ TITLE OF INVENTION: TNF Receptors, TNF Binding Proteins and DNAs Coding for
/ FILE REFERENCE: 98-385-I
/ CURRENT APPLICATION NUMBER: US/09/898,234
/ CURRENT FILING DATE: 2001-07-03
/ PRIOR APPLICATION NUMBER: 09/525,998
/ PRIOR FILING DATE: 2000-03-15
/ PRIOR APPLICATION NUMBER: 08/383,676
/ PRIOR FILING DATE: 1995-02-01
/ PRIOR APPLICATION NUMBER: 08/153,287
/ PRIOR FILING DATE: 1993-11-17
/ PRIOR APPLICATION NUMBER: 07/821,750
/ PRIOR FILING DATE: 1992-01-02
/ PRIOR APPLICATION NUMBER: 07/511,430
/ PRIOR FILING DATE: 1990-04-20
/ NUMBER OF SEQ ID NOS: 87
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 14
/ LENGTH: 2173
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (245)..(1630)
/ OTHER INFORMATION: Description of Artificial Sequence: ratNF-R8
US-09-898-234-14

Query Match 39.6%; Score 22.6; DB 9; Length 2173;
Best Local Similarity 64.2%; Pred. No. 33;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 5 AAGACTGCATGAATATGTCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
Db 810 AGGAATGTAATGAAGCTGTGCTTACCTCCAGTTGCAATGTCACAAACCCCCAG 862

RESULT 9
US-09-899-429A-24
/ Sequence 24, Application US/09899429A
/ Patent No. US20020169119A1
/ GENERAL INFORMATION:
/ APPLICANT: Hauptmann, Rudolph
/ APPLICANT: Himmler, Adolph
/ APPLICANT: Maurer-Fogy, Ingrid
/ APPLICANT: Stratowa, Christian
/ TITLE OF INVENTION: TNF Receptors, TNF Binding Proteins and DNAs Coding for
/ FILE REFERENCE: 98-385-J
/ CURRENT APPLICATION NUMBER: US/09/899,429A
/ CURRENT FILING DATE: 2001-07-03
/ PRIOR APPLICATION NUMBER: 09/792,356
/ PRIOR FILING DATE: 2000-02-23
/ PRIOR APPLICATION NUMBER: 08/477,639
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: 08/383,676
/ PRIOR FILING DATE: 1995-02-01
/ PRIOR APPLICATION NUMBER: 08/153,287
/ PRIOR FILING DATE: 1993-11-17
/ PRIOR APPLICATION NUMBER: 07/821,750
/ PRIOR FILING DATE: 1992-01-02
/ PRIOR APPLICATION NUMBER: 07/511,430
/ PRIOR FILING DATE: 1990-04-20
/ NUMBER OF SEQ ID NOS: 97
/ SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 24
; LENGTH: 2173
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: ratNF-R8
; NAME/KEY: CDS
; LOCATION: (245)..(1627)
; OS: 09-899-429A-24

Query Match      39.6%; Score 22.6; DB 9; Length 2173;
Best Local Similarity 64.2%; Pred. No. 33;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY  5 AAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   810 AGGAATGATGAAGCTGTGCCTACCTCCAGTTGCAAAATGTCACAAACCCCGAG 862

RESULT 10
US-09-792-356-14
; Sequence 14, Application US/09792356
; Publication No. US20020183485A1
; GENERAL INFORMATION:
; APPLICANT: Hauptmann, Rudolph
; APPLICANT: Himmler, Adolph
; APPLICANT: Maurer-Fogy, Ingrid
; APPLICANT: Stratowa, Christian
; TITLE OF INVENTION: TNF Receptors, TNF Binding Proteins and DNAs Coding for
; FILE REFERENCE: 98,385-G
; CURRENT APPLICATION NUMBER: US/09/792,356
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 08/477,639
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/383,676
; PRIOR FILING DATE: 1995-02-01
; PRIOR APPLICATION NUMBER: 08/153,287
; PRIOR FILING DATE: 1993-11-17
; PRIOR APPLICATION NUMBER: 07/821,750
; PRIOR FILING DATE: 1992-01-02
; PRIOR APPLICATION NUMBER: 07/511,430
; PRIOR FILING DATE: 1990-04-20
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2173
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (245)..(1630)
; OTHER INFORMATION: Description of Artificial Sequence: ratNF-R8
US-09-899-422-14

Query Match      39.6%; Score 22.6; DB 10; Length 2173;
Best Local Similarity 64.2%; Pred. No. 33;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY  5 AAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   810 AGGAATGATGAAGCTGTGCCTACCTCCAGTTGCAAAATGTCACAAACCCCGAG 862

RESULT 11
US-09-792-356-14
; Sequence 14, Application US/09792356
; Publication No. US20020183485A1
; GENERAL INFORMATION:
; APPLICANT: Hauptmann, Rudolph
; APPLICANT: Himmler, Adolph
; APPLICANT: Maurer-Fogy, Ingrid
; APPLICANT: Stratowa, Christian
; TITLE OF INVENTION: TNF Receptors, TNF Binding Proteins and DNAs Coding for
; FILE REFERENCE: 98,385-H
; CURRENT APPLICATION NUMBER: US/09/899,422
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 09/525,998
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 08/383,676
; PRIOR FILING DATE: 1995-02-01
; PRIOR APPLICATION NUMBER: 08/153,287
; PRIOR FILING DATE: 1993-11-17
; PRIOR APPLICATION NUMBER: 07/821,750
; PRIOR FILING DATE: 1992-01-02
; PRIOR APPLICATION NUMBER: 07/511,430
; PRIOR FILING DATE: 1990-04-20
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2173
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (245)..(1630)
; OTHER INFORMATION: Description of Artificial Sequence: ratNF-R8
US-09-899-422-14

Query Match      39.6%; Score 22.6; DB 9; Length 2173;
Best Local Similarity 64.2%; Pred. No. 33;
Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY  5 AAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   810 AGGAATGATGAAGCTGTGCCTACCTCCAGTTGCAAAATGTCACAAACCCCGAG 862

RESULT 12
US-09-784-554B-3/C
; Sequence 3, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schalein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Query Match      38.6%; Score 22; DB 9; Length 4056;
Best Local Similarity 63.0%; Pred. No. 70;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY  4 TAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   2227 TACCACCGAGCGTTTCCTTAAATACAGCCGTTTCCAAAGTACCAGCTCTCCAG 2174

RESULT 13
US-09-833-381-1348/C
; Sequence 1348, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
```

; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1348
; LENGTH: 653
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(653)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1348

Query Match 38.2%; Score 21.8; DB 10; Length 653;
Best Local Similarity 65.3%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 8 ACTGCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCA 56
DB 391 AATCAGATAGGATCACACACGAAGTGCATGACATGCCATCAA 343

RESULT 14

US-09-974-300-3301/c
; Sequence 3301, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3301
; LENGTH: 373
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-3301

Query Match 37.9%; Score 21.6; DB 10; Length 373;
Best Local Similarity 68.2%; Pred. No. 45;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATC 54
DB 350 GTATGCATACTTCAACATCAGTTACCATGGAAGGCGATC 307

RESULT 15

US-09-864-761-13274
; Sequence 13274, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 13274
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138498.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-13274

Query Match 37.9%; Score 21.6; DB 10; Length 559;
Best Local Similarity 68.2%; Pred. No. 51;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 13 ATGAATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCA 56
DB 243 ATACAAAGTGAACAGCCCTATTACCAATTTTAGGGTGAACAA 286

Search completed: April 16, 2003, 01:14:17
Job time : 11.6955 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:26:41 ; Search time 15.2104 Seconds
(without alignments)
8439.207 Million cell updates/sec

Title: US-09-647-019-7

Perfect score: 57

Sequence: 1 ggataagactgcataatat.....ccaatgttagccatccag 57

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	428	21	AAC01483
2	57	100.0	887	20	AAK090904
3	55.4	97.2	886	24	AAK090904
4	42.6	77.7	778	20	AAK090903
5	26	45.6	3098	22	AAK29924
6	25.4	44.6	210	19	AAK12902
7	25.4	44.6	210	19	AAK11743
8	25.4	44.6	7677	23	ABV23205
9	25.4	44.6	7677	23	ABV29042
					Human secreted pro
					CDNA encoding huma
					Human 66214 EST cl
					CDNA encoding muri
					C albicans apoptos
					Human biallelic po
					Human biallelic po
					Human prostate exp
					Human prostate exp

C	10	24	42.1	711	21	AAK15033
	11	24	42.1	20689	22	AAK78366
	12	24	42.1	31314	22	AAK37510
C	13	24	42.1	103929	21	AAF22287
	14	23.8	41.8	219	19	AAV13779
C	15	23.6	41.4	8661	23	ABL16184
C	16	23.6	41.4	15320	23	ABL07834
C	17	23.6	41.4	611590	21	AAF22303
	18	23.4	41.1	408	23	ABV18116
	19	23.4	41.1	497	23	ABV47904
C	20	23.4	41.1	9262	23	ABL27260
C	21	23.4	41.1	12718	23	ABL27258
C	22	23.2	40.7	533	22	AAK39818
C	23	23.2	40.7	533	22	AAK90174
C	24	23.2	40.7	954	22	AAK32330
C	25	23.2	40.7	957	22	AAK42259
C	26	23.2	40.7	957	24	ABK68605
C	27	23.2	40.7	957	24	ABK37545
C	28	23.2	40.7	980	24	ABQ88355
C	29	23.2	40.7	990	24	ABQ88354
C	30	23.2	40.7	1539	22	AAK59816
C	31	23.2	40.7	1545	22	AAK59859
C	32	23.2	40.7	1826	22	AAK19580
C	33	23.2	40.7	5698	22	AAK32873
C	34	23.2	40.7	5703	22	AAK32874
C	35	23.2	40.7	18402	21	AAK93705
C	36	23.2	40.7	32222	22	AAK34734
C	37	23	40.4	398	24	ABN65201
*	38	23	40.4	3690	23	ABL29152
C	39	23	40.4	7043	20	AAK13225
C	40	22.8	39.6	428	22	AAK0850
	41	22.6	39.6	432	22	AAK04321
	42	22.6	39.6	432	22	AAK04322
	43	22.6	39.6	2130	24	ABK63694
	44	22.6	39.6	2173	11	AAQ06284
	45	22.6	39.6			

ALIGNMENTS

RESULT 1

AAC01483

ID AAC01483 standard; cDNA; 428 BP.

XX AAC01483;

XX 06-OCT-2000 (first entry)

DT Human secreted protein 5' EST, SEQ ID NO: 1481.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
DE Gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

XX 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

XX P-PSDB; AAG01477.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

Trichoderma reesei
Human immune/haema
Human musculooskele
BAC containing rep
Hepatitis G virus
Drosophila melanog
Drosophila melanog
Arabidopsis thalia
Human prostate exp
Human prostate exp
Drosophila melanog
Drosophila melanog
Genomic sequence #
Human digestive sy
Human olfactory re
Human cDNA encodin
Human DNA for olfa
DNA encoding G-cou
Human G protein co
Human G protein co
Human novel cytol
Human novel cytol
Human G-protein co
Human genomic DNA
Human genomic DNA
CYB5RP fatty acid
Human DNA for a no
Human cancer relat
Drosophila melanog
Drosophila melanog
Enterococcus faeca
Human reproductive
Human reproductive
Human reproductive
Rat sequence diffe
Rat Tumour Necrosi

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 1481; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

CC vectors.

XX SQ Sequence 428 BP; 123 A; 111 C; 104 G; 90 T; 0 other;

Query Match 100.0%; Score 57; DB 21; Length 428;

Best Local Similarity 100.0%; Pred. No. 2.4e-12;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG-57

DB 236 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 292

RESULT 2

AAAX90904

ID AAX90904 standard; cDNA; 887 BP.

AC AAX90904;

XX 17-JAN-2000 (first entry)

XX cDNA encoding human chisel (Csl) gene.

XX Chisel gene; Csl; EF-Hand protein super family; muscle development;

KW heart/skeletal muscle cell development; signalling pathway; regulation;

KW Xq21.3-q22; adaptive process; muscle homeostasis; skeletal myopathy;

KW detection; diagnosis; prophylaxis; treatment; cardiac hypertrophy;

KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;

KW Becker's myotonic dystrophy; heart failure; differentiation; exon;

KW gene therapy; transgenic animal; drug screening; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT exon 1..172

FT /*tag= a

FT /label= Exon_1

FT /note= "Corresponds to residues 19497-19327 of

FT human cosmid clone U228D4"

FT 173..229

FT /*tag= b

FT /label= Exon_2

FT /note= "Corresponds to residues 15687-15631 of

FT human cosmid clone U228D4"

FT 185..451

FT /*tag= c

FT /product= "Chisel (Csl) protein"

FT /note= "Expressed predominantly in heart muscles"

FT 230..316

FT /*tag= d

FT /label= Exon_3

FT /note= "Corresponds to residues 5220-5134 of

FT human cosmid clone U228D4"

FT 317..465

FT /*tag= e

FT /label= Exon_4

FT FT /note= "Corresponds to residues 35384-35236 of

FT human cosmid clone U112E8"

FT 456..887

FT /*tag= f

FT /label= Exon_5

FT /note= "Corresponds to residues 4101-3680 of

FT human cosmid clone U112E8"

XX WO9950410-A1.

XX 07-OCT-1999.

XX 26-MAR-1999; 99WO-AU00220.

XX 27-MAR-1998; 98AU-0002634.

XX (CHAN-) CHANG CARDIAC RES INST VICTOR.

XX (GEHO) GEN HOSPITAL CORP.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;

XX WPI; 1999-610852/52.

XX P-PSDB; AAY28651.

XX Isolated nucleic acids encoding chisel, used to develop products for

PT treating cardiomyopathy, cardiac hypertrophy, heart failure and

PT muscular myopathies -

XX Claim 8; Page 149-150; 157pp; English.

XX The present sequence is the cDNA encoding the human chisel gene (Csl)

CC that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.

CC It is a member of the EF-Hand protein super family and is involved in

CC signalling pathways. It is predominantly expressed in heart and skeletal

CC muscles and is activated after the differentiation of cells. Csl

CC functions in regulation aspects of differentiation or adaptive processes

CC that maintain muscle homeostasis. This sequence can be used in the

CC detection, diagnosis, prophylactic and therapeutic treatment of diseases

CC such as those involving aberrant muscle cell development and functional

CC activity. It is also used in the treatment of muscular and myotonic

CC dystrophies, skeletal myopathies such as Duchenne muscular dystrophy and

CC Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,

CC myocarditis, myofiber atrophy, etc. The Csl gene sequence can also be

CC used in gene therapy, for the production of transgenic animals and for

CC drug screening.

XX SQ Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;

Query Match 100.0%; Score 57; DB 20; Length 887;

Best Local Similarity 100.0%; Pred. No. 2.8e-12;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57

DB 173 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 229

RESULT 3

AAAD27216

ID AAD27216 standard; DNA; 886 BP.

XX AAD27216;

XX 09-APR-2002 (first entry)

XX Human 66214 EST clone DNA.

XX Human; congestive heart failure; dilative cardiomyopathy; sudden death;

KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;

KW heart muscle disease; conduction disorder; coronary heart disease;

KW systemic arterial hypertension; pulmonary hypertension; endocarditis;

KW pulmonary heart disease; valvular heart disease; pericardial disease;

KW congenital heart disease; gene therapy; syncope; transgenic animal;
 XX expressed sequence tag; EST; clone 66214; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 184..450
 FT /*tag= a
 FT /product= "Human 66214 protein"
 FT misc_feature 298..588
 FT /*tag= b
 FT /note= "66214 cDNA fragment"
 FT polyA_signal 857..862
 FT /*tag= c
 XX WO200192567-A2.
 PD 06-DEC-2001.
 XX 30-MAY-2001; 2001WO-EP06165.
 XX 30-MAY-2000; 2000US-207400P.
 XX (MEDI-) MEDIGENE AG.
 XX Bunk D, Reuner B, Beck J, Henkel T;
 XX WPI; 2002-122073/16.
 XX P-PSDB; AAJ16632.
 XX Identifying a subject at risk for a heart disease e.g. congestive heart
 PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 PT the polypeptide expressed by genes abnormally expressed in heart tissue
 PT -
 XX Claim 2a; Fig 9b; 154pp; English.
 PS The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present DNA sequence is expressed sequence tag
 CC (EST) 66214 clone.
 XX SQ Sequence 886 BP; 278 A; 172 C; 191 G; 245 T; 0 other;
 Query Match 97.2%; Score 55.4; DB 24; Length 886;
 Best Local Similarity 98.2%; Pred. No. 1.2e-11;
 Matches 56; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57
 |||||
 Db 172 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 228
 |||||
 RESULT 4
 AAX90903
 ID AAX90903 standard; cDNA; 778 BP.
 XX AAX90903;
 AC AAX90903;
 XX 17-JAN-2000 (first entry)
 DT
 XX

DE cDNA encoding murine chisel (Csl) gene.
 XX Chisel gene; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; murine;
 KW X chromosome; regulation; adaptive process; muscle homeostasis;
 KW detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KW differentiation; gene therapy; transgenic animal; drug screening; es.
 XX Mus sp.
 OS
 XX Key Location/Qualifiers
 FH 199..456
 FT CDS
 FT /*tag= a
 FT /product= "chisel (Csl) protein"
 FT /note= "Expressed especially in heart muscles"
 XX WO9950410-A1.
 XX 07-OCT-1999.
 XX 26-MAR-1999; 99WO-AU00220.
 XX 27-MAR-1998; 98AU-0002634.
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 XX (GEHO) GEN HOSPITAL CORP.
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX WPI; 1999-610852/52.
 XX P-PSDB; AAY28650.
 XX Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX Claim 3; Page 148; 157pp; English.
 XX The present sequence is the cDNA encoding the murine chisel gene (Csl)
 CC that is mapped to the mouse X chromosome. It is a member of the EF-Hand
 CC protein super family and is involved in signalling pathways. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX SQ Sequence 778 BP; 231 A; 166 C; 179 G; 202 T; 0 other;
 Query Match 74.7%; Score 42.6; DB 20; Length 778;
 Best Local Similarity 84.2%; Pred. No. 1e-06;
 Matches 48; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57
 |||||
 Db 181 GGCTAAGACCTTGTAATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 237
 |||||
 RESULT 5
 AAH29924
 ID AAH29924 standard; DNA; 3098 BP.
 XX AAH29924;
 AC AAH29924;
 XX

DT 27-JUL-2001 (first entry)
 XX C albicans apoptosis associated coding sequence #68.
 DE
 XX
 XX Yeast, fungus; apoptosis; infection; proliferative disease;
 KW vaccine; autoimmune disease; ischaemia; neurodegeneration; ds.
 XX
 OS Candida albicans.
 XX
 XX WO200102550-A2.
 PN
 XX 11-JAN-2001.
 PD
 XX
 XX 03-JUL-2000; 2000WO-BE00077.
 PF
 XX
 XX 01-JUL-1999; 99EP-0870141.
 PR
 XX (JANC) JANSSEN PHARM NV.
 PA
 XX Contreras RH, De Backer MD, Luyten-WHML, Malcorps IKL;
 PI Nelissen BJM, Reekmans RJ;
 XX WPI; 2001-367042/38.
 DR P-PSDB; AAG70888.
 XX
 XX Yeast and fungal nucleic acids encoding proteins involved in a pathway
 PT leading to programmed cell death, useful for treating proliferative
 PT disorders, yeast and fungal infections, or for preventing apoptosis in
 PT certain diseases -
 XX
 XX Claim 17; Fig 2; 218pp; English.
 PS
 XX The present invention provides the protein and coding sequences of a
 CC number of apoptosis associated proteins from the yeast *Saccharomyces*
 CC cerevisiae and the fungus *Candida albicans*. These can be used to identify
 CC treatments for fungal and yeast infections, for proliferative diseases
 CC and for apoptosis related diseases such as autoimmune diseases, ischaemia
 CC and neurodegeneration. The present sequence is one of the *C. albicans*
 CC coding sequences of the invention.
 XX
 XX Sequence 3098 BP; 1023 A; 528 C; 539 G; 1008 T; 0 other;
 SQ
 Query Match 45.6%; Score 26; DB 22; Length 3098;
 Best Local Similarity 76.2%; Pred. No. 3.7;
 Matches 32; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 6 AGACTGATGATATGTCGAACAGCCAGTTTCCAAATGTTAG 47
 DB 2459 AGATTCTGAATTGTACAAACCCGCGCAGTTTCAATTTTGG 2500
 RESULT 6
 AAX12902
 ID AAX12902 standard; DNA; 210 BP.
 AC
 AC AAX12902;
 XX
 XX 30-MAR-1999 (first entry)
 DT
 DE Human biallelic polymorphic DNA fragment EST398528.
 XX
 XX Polymorphism; biallelic; human; forensic; paternity testing; disease;
 KW detection; phenotypic typing; characteristic; infection; hereditary;
 KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
 KW treatment; marker; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO9820165-A2.
 PN
 XX 14-MAY-1998.
 PD
 XX 05-NOV-1997; 97WO-US20313.
 PF

XX 06-NOV-1996; 96US-0030455.
 PR (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX Hudson T, Lander ES, Wang D;
 PI
 XX WPI; 1998-286974/25.
 DR
 XX New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics, paternity
 PT testing or phenotypic typing for disease
 XX
 XX Claim 1; Page 301; 310pp; English.
 PS
 XX AAX10269-X12937 are human DNA fragments which contain biallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
 CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.
 XX
 SQ Sequence 210 BP; 56 A; 53 C; 50 G; 51 T; 0 other;
 Query Match 44.6%; Score 25.4; DB 19; Length 210;
 Best Local Similarity 74.4%; Pred. No. 3.4;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 8 ACTGCATGATATGTCGAACAGCCAGTTTCCAAATGTTAGGC 50
 DB 130 ACAGCAGGAACACGTGGAAAGGCGCTTTCCAGTGTAAAGGC 172
 RESULT 7
 AAX11743
 ID AAX11743 standard; DNA; 210 BP.
 AC
 AC AAX11743;
 XX
 XX 30-MAR-1999 (first entry)
 DT
 DE Human biallelic polymorphic DNA fragment EST398528.
 XX
 XX Polymorphism; biallelic; human; forensic; paternity testing; disease;
 KW detection; phenotypic typing; characteristic; infection; hereditary;
 KW autoimmune disease; cancer; inflammation; drug; therapy; medicament;
 KW treatment; marker; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO9820165-A2.
 PN
 XX 14-MAY-1998.
 PD
 XX 05-NOV-1997; 97WO-US20313.
 PF
 XX 06-NOV-1996; 96US-0030455.
 PR (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX

PI Hudson T, Lander ES, Wang D;
 XX
 DR WPI; 1998-286974/25.
 XX
 PT New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics, paternity
 PT testing or phenotypic typing for disease
 XX
 XX
 PS Claim 1; Page 189; 310pp; English.
 XX
 CC AAX10269-X12937 are human DNA fragments which contain biallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed,
 CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.
 XX
 SQ Sequence 210 BP; 56 A; 52 C; 50 G; 51 T; 1 other;
 Query Match 44.6%; Score 25.4; DB 19; Length 210;
 Best Local Similarity 74.4%; Pred. No. 3.4;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 OY 8 ACTGCATGATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGC 50
 Db 130 ACAGCAGGACACGTGGAAAGGCTGTTCACAGTGTAGGC 172
 RESULT 8
 ABV23205/c
 ID ABV23205 standard; cDNA; 7677 BP.
 AC ABV23205;
 XX
 XX 16-SEP-2002 (first entry)
 DT
 XX Human prostate expression marker cDNA 23196.
 DE
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 KW Homo sapiens.
 OS
 XX WO200160860-A2.
 PN 23-AUG-2001.
 XX
 XX 20-FEB-2001; 2001WO-US05171.
 XX
 XX 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA
 XX Schlegel R, Endege WO, Monahan JE;
 PI
 XX WPI; 2001-662795/76.

DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 XX Claim 1; Page 4173-4175; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 7677 BP; 2389 A; 1642 C; 1794 G; 1846 T; 6 other;
 Query Match 44.6%; Score 25.4; DB 23; Length 7677;
 Best Local Similarity 74.4%; Pred. No. 7.9;
 Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 OY 10 TGCATGATATGTCGAAACAGCCAGTTCCTCAATGTTAGAGCCA 52
 Db 1412 TGTATGTCCTGTGGCAGTAGCCAGTTTCGATGGCAGAGCCA 1370
 RESULT 9
 ABV29042/c
 ID ABV29042 standard; cDNA; 7677 BP.
 XX
 XX AC ABV29042;
 XX
 XX 16-SEP-2002 (first entry)
 DT
 XX Human prostate expression marker cDNA 29033.
 DE
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 KW Homo sapiens.
 OS
 XX WO200160860-A2.
 PN 23-AUG-2001.
 XX
 XX 20-FEB-2001; 2001WO-US05171.
 XX
 XX 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA
 XX Schlegel R, Endege WO, Monahan JE;
 PI
 XX WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX

PS Claim 1; Page 6152-6154; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in tables 1-9 (ABV0010-ABV62213) of the

CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate

CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer

CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;

CC (h) assessing the aggressiveness or indolence of prostate cancer in a

CC patient;

CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX

SQ Sequence 7677 BP; 2389 A; 1642 C; 1794 G; 1846 T; 6 other;

Query Match 44.6%; Score 25.4; DB 23; Length 7677;

Best Local Similarity 74.4%; Pred. No. 7.9;

Matches 32; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 10 TGCATGAATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCA 52

DB 1412 TGTATGCTCTGTGCGAGTAGCCAGTTCCCATGCGAGAGCCA 1370

RESULT 10

AAFL15033/c

ID AAF15033 standard; CDNA; 711 BP.

XX

AC AAF15033;

XX

DT 13-MAR-2001 (first entry)

XX

DE Trichoderma reesei EST SEQ ID NO:7556.

XX

KW Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX

OS Trichoderma reesei.

XX

PN WO200056762-A2.

XX

PD 28-SEP-2000.

XX

PF 22-MAR-2000; 2000WO-US07781.

XX

PR 22-MAR-1999; 99US-0273623.

XX

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

XX

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX

DR WPI; 2000-594572/56.

XX

PT Monitoring differential expression of genes in filamentous fungal cells

PT uses fluorescence-labeled nucleic acids isolated from the cells and a

PT substrate of expressed sequence tags -

XX

PS Claim 89; Page 3052; 3161pp; English.

XX

CC The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from

CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs

CC

CC are used in the methods for monitoring differential expression of genes

CC in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from FF cells allows the production

CC potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be

CC identified and gene copy number variation and stability can be

CC monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore

CC morphogenesis, recombination, metabolic or catabolic pathway

CC engineering. Using ESTs provides several advantages over genomic or

CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the

CC microarrays based on function of the gene products to facilitate

CC analysis of the results. AAF07478 to AAF11247 represents ESTs from

CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus

CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are

CC all specifically claimed in the present invention.

XX

SQ Sequence 711 BP; 133 A; 224 C; 188 G; 153 T; 13 other;

Query Match 42.1%; Score 24; DB 21; Length 711;

Best Local Similarity 68.8%; Pred. No. 16;

Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6 AGACTGATGAATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCAT 53

DB 109 AAATGCTTGTGAAGTGTTCGAGAGCCGGTTCCCAATTCAGAGCGAT 62

RESULT 11

AAK78366

ID AAK78366 standard; DNA; 20689 BP.

XX

AC AAK78366;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33178.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN WO200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 33178; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cystostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX

SQL	Sequence	20689 BP; 4818 A; 5626 C; 5868 G; 4377 T; 0 other;
	Query Match	42.1%; Score 24; DB 22; Length 20689;
	Best Local Similarity	68.8%; Pred. No. 35;
	Matches	33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY	2	GATAGACTGCATGATATGTGGAACAGCCAGTTTCCAATTGATAGAG 49
Dd	5772	GAGAAGCTTAGCTTGAAAGCTGCGAACCACCACTTCAGAGCTAGAG 5819
RESULT 12		
AAL37510		
ID	AAL37510 standard; DNA; 31314 BP.	
XX	XX	
AC	AAL37510;	
XX	XX	
DT	08-JAN-2002 (first entry)	
XX	XX	
DE	Human musculoskeletal system related polynucleotide SEQ ID NO 3875.	
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;	
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KW	neurological disease; infection; human; secreted protein;	
KW	musculoskeletal system; ds.	
XX	XX	
OS	Homo sapiens.	
XX	XX	
PN	WO200155367-A1.	
XX	XX	
PD	02-AUG-2001.	
XX	XX	
PF	17-JAN-2001; 2001WO-US01338.	
XX	XX	
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0198874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226868.	
PR	23-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227009.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	01-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	
PR	05-SEP-2000; 2000US-0229513.	
PR	06-SEP-2000; 2000US-0230437.	
PR	06-SEP-2000; 2000US-0230438.	
PR	08-SEP-2000; 2000US-0231242.	
PR	08-SEP-2000; 2000US-0231243.	
PR	08-SEP-2000; 2000US-0231244.	
PR	08-SEP-2000; 2000US-0231413.	
PR	08-SEP-2000; 2000US-0231414.	
PR	08-SEP-2000; 2000US-0232080.	
PR	08-SEP-2000; 2000US-0232081.	
PR	12-SEP-2000; 2000US-0231968.	
PR	14-SEP-2000; 2000US-0232397.	
PR	14-SEP-2000; 2000US-0232398.	
PR	14-SEP-2000; 2000US-0232399.	
PR	14-SEP-2000; 2000US-0232400.	
PR	14-SEP-2000; 2000US-0232401.	
PR	14-SEP-2000; 2000US-0233063.	
PR	14-SEP-2000; 2000US-0233064.	
PR	14-SEP-2000; 2000US-0233065.	
PR	21-SEP-2000; 2000US-0234273.	
PR	21-SEP-2000; 2000US-0234274.	
PR	25-SEP-2000; 2000US-0234997.	
PR	25-SEP-2000; 2000US-0234998.	
PR	26-SEP-2000; 2000US-0235484.	
PR	27-SEP-2000; 2000US-0235834.	
PR	27-SEP-2000; 2000US-0235836.	
PR	29-SEP-2000; 2000US-0236327.	
PR	29-SEP-2000; 2000US-0236367.	
PR	29-SEP-2000; 2000US-0236368.	
PR	29-SEP-2000; 2000US-0236369.	
PR	29-SEP-2000; 2000US-0236370.	
PR	02-OCT-2000; 2000US-0236802.	

PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-451937/48.

XX Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -

XX Example 2; SEQ ID NO 3875; 781pp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 31314 BP; 10461 A; 5921 C; 5585 G; 9347 T; 0 other;

Query Match 42.1%; Score 24; DB 22; Length 31314;
 Best Local Similarity 68.8%; Pred. No. 38;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 5 AAGACTGCATGATATGTCGAAACAGCCAGTTTCCATGTTAGAGCCA 52

Db 1967 AATCTGCAAAATTTGGTGAAAGAGAGTCTGCAGTGTTAAGCCA 2014

RESULT 13

AAF22287/c

ID AAF22287 standard; DNA; 103929 BP.
 XX AAF22287;
 AC AAF22287;
 XX 20-MAR-2001 (first entry)
 DT 20-MAR-2001 (first entry)
 XX BAC containing repeats from centromeres 1-4 #10.
 DE BAC containing repeats from centromeres 1-4 #10.
 XX Centromere; microsome; vector; ds.
 KW Centromere; microsome; vector; ds.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO2000055325-A2.
 PN WO2000055325-A2.
 XX 21-SEP-2000.
 PD 21-SEP-2000.
 XX 17-MAR-2000; 2000WO-US07392.
 PF 17-MAR-2000; 2000WO-US07392.
 XX 18-MAR-1999; 99US-0125219.
 PR 01-APR-1999; 99US-0127409.
 PR 18-MAY-1999; 99US-0134770.
 PR 13-SEP-1999; 99US-0153584.
 PR 17-SEP-1999; 99US-0154603.
 XX (UYCH-) UNIV CHICAGO.
 PA (UYCH-) UNIV CHICAGO.
 XX Preuss D, Copenhaver G, Keith K;
 PI Preuss D, Copenhaver G, Keith K;
 XX WPI; 2000-587529/55.
 DR WPI; 2000-587529/55.
 XX Recombinant DNA construct comprising a plant centromere, useful for
 PT producing stably inherited microsome which can serve as vectors for
 PT the construction of transgenic plant and animal cells -
 XX Claim 102; Page 484-508; 1449pp; English.
 PS Claim 102; Page 484-508; 1449pp; English.
 XX The present invention relates to a recombinant DNA construct of a plant
 CC (Arabidopsis thaliana) centromere. The constructs are useful for
 CC producing stably inherited microsome which can serve as vectors for
 CC the construction of transgenic plant and animal cells expressing
 CC selected proteins such as hormones, enzymes, interleukins, clotting
 CC factors, cytokines, antibodies, and growth factors.
 XX Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33085 T; 0 other;
 SQ Sequence 103929 BP; 32987 A; 19310 C; 18547 G; 33085 T; 0 other;
 Query Match 42.1%; Score 24; DB 21; Length 103929;
 Best Local Similarity 64.3%; Pred. No. 51;
 Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 OY 2 GATAAGCTGCATGATATGTCGAAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
 Db 99535 GGTTAAGTGCATGATATGTCGAAACAGCCAGTATTACGGTTTAAATGTTATCCAG 99480
 RESULT 14
 AAV13779
 ID AAV13779 standard; DNA; 219 BP.
 XX AAV13779;
 AC AAV13779;
 XX 08-MAY-1998 (first entry)
 DT 08-MAY-1998 (first entry)
 XX Hepatitis G virus nucleic acid fragment.
 DE Hepatitis G virus nucleic acid fragment.
 XX PCR primer; detection; non-A-B-C-D-E-F type hepatitis virus;
 KW PCR primer; detection; non-A-B-C-D-E-F type hepatitis virus;
 XX HGV; ss.
 XX Hepatitis G virus.
 OS Hepatitis G virus.
 XX JP10000092-A.
 PN JP10000092-A.
 XX 06-JAN-1998.
 PD 06-JAN-1998.

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 ; Search time 129.143 Seconds
(without alignments)
12845.099 Million cell updates/sec

Title: US-09-647-019-7
Perfect score: 57
Sequence: 1 ggataagactgatgaatat.....ccaatgttagacccatccag 57

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_mu.*
- 22: em_mu.*
- 23: em_mu.*
- 24: em_mu.*
- 25: em_mu.*
- 26: em_mu.*
- 27: em_mu.*
- 28: em_mu.*
- 29: em_mu.*
- 30: em_mu.*
- 31: em_mu.*
- 32: em_mu.*
- 33: em_mu.*
- 34: em_mu.*
- 35: em_mu.*
- 36: em_mu.*
- 37: em_mu.*
- 38: em_mu.*
- 39: em_mu.*
- 40: em_mu.*
- 41: em_mu.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	57	100.0	835	9	BC005948	Homo sapi
2	57	100.0	885	9	HS250584	Homo sapi
3	57	100.0	886	9	AF129505	Homo sapi
C 4	57	100.0	47440	9	U73509	Homo sapien
C 5	57	100.0	150319	2	AL772370	Homo sapi
C 6	55.4	97.2	886	6	AX322774	Sequence
7	42.6	74.7	787	10	AY026524	Mus muscu
8	42.6	74.7	936	10	AF364070	Mus muscu
9	42.6	74.7	943	10	MMU245772	Mus muscu
10	39.4	69.1	892	10	AF364071	Mus muscu
11	29.8	52.3	230579	2	AC124177	Rattus no
C 12	26.8	47.0	44942	9	AL589691	Human DNA
C 13	26.8	47.0	168235	2	AL357505	Homo sapi
C 14	26.6	46.7	375	4	D26519	Muscula zib
C 15	26.6	46.7	401	4	L77957	Martea zib
C 16	26.6	46.7	402	4	AB029420	Martea zib
C 17	26.6	46.7	402	4	AB029421	Martea zib
C 18	26.6	46.7	402	4	AB029423	Martea zib
C 19	26.6	46.7	1140	4	AB012356	Martea zib
C 20	26.6	46.7	1140	4	AB012358	Martea zib
C 21	26.6	46.7	1140	4	AB012359	Martea zib
C 22	26.6	46.7	1140	4	AB012360	Martea zib
C 23	26.6	46.7	1140	4	AB012361	Martea zib
C 24	26.6	46.7	1140	4	AF448241	Martea zib
C 25	26.6	46.7	1140	4	AF448242	Martea zib
C 26	26.6	46.7	1140	4	AF448243	Martea zib
C 27	26.6	46.7	1140	4	AF448244	Martea zib
C 28	26.6	46.7	37635	9	AC004152	Homo sapi
C 29	26.2	46.0	923	5	AF343894	Xenopus l
30	26	45.6	2598	6	AX489570	Sequence
31	26	45.6	3098	6	AX073292	Sequence
C 32	26	45.6	113686	9	AC060834	Homo sapi
C 33	25.6	44.9	34498	10	AF321235	Mus muscu
C 34	25.6	44.9	66142	2	AC119274	Mus muscu
C 35	25.6	44.9	118684	9	AC004259	Human Chr
C 36	25.6	44.9	133863	9	AC004600	Homo sapi
C 37	25.6	44.9	155781	2	AF321234	Mus muscu
C 38	25.6	44.9	183426	9	AC100774	Homo sapi
C 39	25.6	44.9	184595	2	AC084009	Homo sapi
C 40	25.6	44.9	190749	2	AF321233	Mus muscu
C 41	25.6	44.9	195974	2	AC122258	Mus muscu
C 42	25.4	44.6	210	11	G44373	WIAF-4237-S
C 43	25.4	44.6	6401	9	HSU09042	Human myob
C 44	25.4	44.6	6401	9	HSY07759	H.sapiens m
C 45	25.4	44.6	14206	9	HSAT3	H.sapiens g

ALIGNMENTS

RESULT 1
BC005948
LOCUS
DEFINITION
Homo sapiens, small muscle protein, X-linked, clone MGC:14584
IMAGE:4246501, mRNA, complete cds.
ACCESSION
BC005948
VERSION
BC005948.1 GI:13543590
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 835)
AUTHORS
Straussberg, R.
TITLE
Direct Submission

JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT NTH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-x@mail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 21 Row: a Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6625646.

FEATURES
 Location/Qualifiers
 1..835
 /organism="Homo sapiens"
 /db_xref="LocusID:23676"
 /db_xref="taxon:9606"
 /clone="MGC:14584 IMAGE:4246501"
 /tissue type="Skeletal Muscle"
 /clone_lib="NIH MGC_81"
 /lab_host="DH10B"
 /notes="Vector: pDNR-LIB"
 113..379
 /codon_start=1
 /product="small muscle protein, X-linked"
 /protein_id="AAH05948.1"
 /db_xref="GI:13543591"
 /translation="NMMSKQPVSNVRAIQANINIPMGAFRPGAGQPPRRKECTPEVEE
 GVPTSDDEKKPIPGAKLPGPAVNLSIQNIKSELKYVPKAEQ"

BASE COUNT 282 a 155 c 171 g 227 t

ORIGIN

Query Match 100.0%; Score 57; DB 9; Length 835;
 Best Local Similarity 100.0%; Pred. No. 4.6e-11;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
 |||||
 Db 101 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 157

RESULT 2
HS250584
LOCUS Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
DEFINITION protein (Srmx gene).
ACCESSION AJ250584
VERSION AJ250584.1 GI:10178976
KEYWORDS Srmx gene; stretch responsive muscle (X-chromosome).
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 885)

TITLE Kemp.T.J., Sadusky.T.J., Simon.M., Brown.R., Eastwood.M., Sassoon.D.A. and Coulton.G.R.
 Identification of a novel stretch-responsive skeletal muscle gene (Smpx)

JOURNAL Genomics 72 (3), 260-271 (2001)
MEDLINE 21295047
PUBMED 11401441
REFERENCE 2 (bases 1 to 885)
AUTHORS Kemp.T.J.

TITLE Direct Submission
JOURNAL Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial College of Science Technology and Medicine, SAF Bldg, Level 2, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
 Related sequences: AJ245772, U73508 to U73509.

COMMENT
FEATURES
 Location/Qualifiers
 1..885
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="skeletal muscle"
 1..885
 /genes="Srmx"
 1..183
 /gene="Srmx"
 184..450
 /gene="Srmx"
 /codon_start=1
 /product="stretch responsive muscle (X-chromosome)"
 /protein_id="CAC08492.1"
 /db_xref="GI:10178977"
 /translation="NMMSKQPVSNVRAIQANINIPMGAFRPGAGQPPRRKECTPEVEE
 GVPPTSDDEKKPIPGAKLPGPAVNLSIQNIKSELKYVPKAEQ"

3'UTR 451..885
 /gene="Srmx"
 857..862
 /gene="Srmx"
 172 c 191 g 244 t

polyA_signal

BASE COUNT 278 a 172 c 191 g 244 t

ORIGIN

Query Match 100.0%; Score 57; DB 9; Length 885;
 Best Local Similarity 100.0%; Pred. No. 4.6e-11;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
 |||||
 Db 172 GGATAAGACTGCATGATATGTCGAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 228

RESULT 3
AF129505
LOCUS Homo sapiens small muscular protein (SMPX) mRNA, complete cds.
DEFINITION Homo sapiens small muscular protein (SMPX) mRNA, complete cds.
ACCESSION AF129505
VERSION AF129505.1 GI:6625646
KEYWORDS

SOURCE Homo sapiens.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 886)

REFERENCE
AUTHORS Patzak.D., Zhuchenko.O., Lee,C.C. and Wehnert,M.
TITLE Identification, mapping, and genomic structure of a novel X-chromosomal human gene (SMPX) encoding a small muscular protein

JOURNAL Hum. Genet. 105 (5), 506-512 (1999)
MEDLINE 20065879
PUBMED 10598820
 2 (bases 1 to 886)

REFERENCE
AUTHORS Patzak.D.
TITLE Direct Submission
JOURNAL Submitted (18-JAN-1999) Molecular Human Genetics, Institut for Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany

FEATURES
 Location/Qualifiers
 1..886
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xq22.1-22.2"
 <1..>886
 /gene="SMPX"

gene <1..>886
5'UTR <1..183
CDS 184..450
 /gene="SMPX"


```

/codon_start=1
/evidence=not experimental
/product="small muscular protein"
/protein_id="AAF19343.1"
/db_xref="GI:6625647"
/translation="MNMSKQPSVNVRAIQANINPMGAFPGAGOPPRRKECTPEVEE
GVPTSDSEKQIPGAKLPGPAVNLSEIQNIKSELKYVPKAEQ"
misc_feature
190
/gene="SMPX"
/notes="alternate position for initiation methionine"
451..>886
/gene="SMPX"
857..862
/polya_signal
/gene="SMPX"
BASE COUNT 279 a 172 c 191 g 244 t
ORIGIN

```

```

Query Match 100.0%; Score 57; DB 9; Length 886;
Best Local Similarity 100.0%; Pred. No. 4.6e-11;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATAGACTGCATGATATGTCGAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
|||||
Db 172 GGATAGACTGCATGATATGTCGAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 228

```

```

RESULT 4
U73509/c
LOCUS Homo sapiens cosmid clone U228D4 from Xp22.1-22.2, complete
DEFINITION sequence.
ACCESSION U73509
VERSION U73509.1 GI:1616809
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston,J.E. and Waterston,R.
1 (bases 1 to 47440)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998) X
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 47440)
AUTHORS Waterston,R.
TITLE The sequence of Homo sapiens cosmid clone U228D4
JOURNAL Unpublished (1999)
REFERENCE 3 (bases 1 to 47440)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-Oct-1996)
REFERENCE 4 (bases 1 to 47440)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1998) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 47440)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (27-APR-1999) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap

```

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

This clone was mapped by Grieff, M., Whyte, M. P., Thakker, R. V., and Mazzarella, R. Sequence analysis of 139 kb in Xp22.1 containing spermine synthase and the 5' region of PEX. Genomics 44:227-231 (1997).

SOURCE INFORMATION:

This clone is from a chromosome X-specific cosmid library LLOXNCC01 'U'. The source of the chromosomes was a human/hamster hybrid, GM07297-F, from Robert Nusbaum at the University of Pennsylvania School of Medicine. Please contact the Lawrence Livermore National Laboratory at <http://www-bio.llnl.gov/genome> to obtain the clone. VECTOR: Lawrist16. Location/Qualifiers

FEATURES

Source

```

1..47440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22.1-22.2"
/clone="U228D4"
/clone_lib="LLOXNCC01-U"

```

repeat_region

```

9..261
/rpt_family="Retroviral"
2020..2180
/rpt_family="MIR"
4824..4862

```

repeat_region

```

/rpt_family="CAAT)n"
complement(5131..5222)
/notes="match to EST N87511 (NID:G1440713)"

```

misc_feature

```

complement(5133..5222)
/notes="match to EST W07478 (NID:G1281479) za96c10.rl"

```

misc_feature

```

complement(5148..5222)
/notes="match to EST AA092554 (NID:G1637327)"

```

misc_feature

```

complement(5169..5222)
/notes="match to EST AA248485 (NID:G1879506)"
7616..7763
/rpt_family="MIR"
8781..8842
/rpt_family="U2"
11721..11940
/rpt_family="MER1_type"
12770..12961
/rpt_family="MIR"
13039..13107
/rpt_family="L2"
13897..14201
/rpt_family="Alu"
14233..14397
/rpt_family="MIR"

```

repeat_region

```

complement(15622..15689)
/notes="match to EST AA248485 (NID:G1879506)"

```

misc_feature

```

complement(15622..15689)
/notes="match to EST W07478 (NID:G1281479) za96c10.rl"
16751..16837
/rpt_family="MIR"

```

repeat_region

```

complement(19314..19497)
/notes="match to EST AA248485 (NID:G1879506)"

```

misc_feature

```

complement(19314..19478)
/notes="match to EST W07478 (NID:G1281479) za96c10.rl"

```

misc_feature

```

complement(19314..19447)
/notes="match to EST AA248067 (NID:G1878536)"

```

misc_feature

```

complement(19314..19423)
/notes="match to EST N87511 (NID:G1440713)"

```

```

misc_feature      complement(19314..19413)
repeat_region     /notes=match to EST AA092554 (NID:g1637327)"/
repeat_region     /rpt_family="Retroviral"
repeat_region     21260..21572
repeat_region     /rpt_family="Alu"
repeat_region     21984..22282
repeat_region     /rpt_family="MIR"
repeat_region     23499..23673
repeat_region     24236..24556
repeat_region     /rpt_family="L1"
repeat_region     24759..24869
repeat_region     /rpt_family="L1"
repeat_region     25920..26860
repeat_region     /rpt_family="L1"
repeat_region     26867..27210
repeat_region     /rpt_family="MaLR"
repeat_region     27213..27810
repeat_region     /rpt_family="L1"
repeat_region     28727..28816
repeat_region     /rpt_family="Retroviral"
repeat_region     28945..29005
repeat_region     /rpt_family="Retroviral"
repeat_region     29715..30007
repeat_region     /rpt_family="Alu"
repeat_region     30773..31151
repeat_region     /rpt_family="MaLR"
repeat_region     31705..33575
repeat_region     /rpt_family="L1"
repeat_region     33587..37330
repeat_region     /rpt_family="L1"
repeat_region     37345..37464
repeat_region     /rpt_family="(TA)n"
repeat_region     37504..37858
repeat_region     /rpt_family="L1"
repeat_region     38102..38146
repeat_region     /rpt_family="MIR"
repeat_region     38578..38706
repeat_region     /rpt_family="L2"
repeat_region     38794..38914
repeat_region     /rpt_family="Retroviral"
repeat_region     38939..39240
repeat_region     /rpt_family="Alu"
repeat_region     39241..39431
repeat_region     /rpt_family="Retroviral"
repeat_region     39469..39659
repeat_region     /rpt_family="MERL_type"
repeat_region     39877..39946
repeat_region     /rpt_family="L2"
repeat_region     40027..40328
repeat_region     /rpt_family="Alu"
repeat_region     42307..42488
repeat_region     /rpt_family="MERL_type"
repeat_region     43876..44349
repeat_region     /rpt_family="MaLR"
repeat_region     45063..45245
repeat_region     /rpt_family="MIR"
repeat_region     45276..45399
repeat_region     /rpt_family="MERL_type"
repeat_region     46393..46517
repeat_region     /rpt_family="L2"
repeat_region     46645..47185
repeat_region     /rpt_family="MERL2_9"
BASE COUNT      15150 a 9644 c 9310 g 13336 t
ORIGIN
Query Match      100.0%; Score 57; DB 9; Length 47440;
Best Local Similarity 100.0%; Pred. No. 3.7e-11;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATAAGACTGCATGATATGTGCAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
Db 15687 GGATAAGACTGCATGATATGTGCAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 15631

```

RESULT 5
AL772370/c
LOCUS
DEFINITION
Homo sapiens chromosome X clone RP11-184B10, *** SEQUENCING IN PROGRESS ***, 2 unordered pieces.
AL772370
VERSION
AL772370.5 GI:22416024
KEYWORDS
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 150319)
AUTHORS
Bird,C.
TITLE
Direct Submission
JOURNAL
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT
On Aug 21, 2002 this sequence version replaced gi:22204612.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA184B10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 99% of reads
Chemistry: Dye-terminator ET-amerham; 0% of reads Consensus
quality: 149793 bases at least Q40
Consensus quality: 149914 bases at least Q30
Consensus quality: 150006 bases at least Q20
Insert size: 150219; sum-of-contigs
Insert size: 152259; 2.5% error; agarose-fp
Quality coverage: 17.18x in Q20 bases; sum-of-contigs Quality
coverage: 17.58x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4859: contig of 4859 bp in length
* 4860 4959: gap of 100 bp
* 4960 150319: contig of 145360 bp in length.

FEATURES
source
1. .150319
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-184B10"
/clone_lib="RPC1-11.1"
1. .4859
misc_feature
/note="assembly fragment:05115
fragment chain:1"
4960..150319
misc_feature
/note="assembly fragment:05270
fragment chain:1
clone end:SP6
vector_side:right"
BASE COUNT 47436 a 29741 c 29722 g 43320 t 100 others
ORIGIN

Query Match 100.0%; Score 57; DB 2; Length 150319;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATAAGACTGCATGATATGTGCAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57

```

Db 28459 GGATAAGACTGTCATGAATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 28403
|||||
RESULT 6
AX322774
LOCUS
DEFINITION Sequence 18 from Patent WO0192567.
ACCESSION AX322774
VERSION AX322774.1 GI:18093754
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Bunk,D., Reuner,B., Beck,J. and Henkel,T.
TITLE Novel target genes for diseases of the heart
JOURNAL Patent: WO 0192567-A 18 06-DEC-2001;
Medigene AG (DE)
FEATURES
Location/Qualifiers
source
1..886
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 278 a 172 c 191 g 245 t
ORIGIN
Query Match 97.2%; Score 55.4; DB 6; Length 886;
Best Local Similarity 98.2%; Pred. No. 1.9e-10;
Matches 56; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGATAAGACTGTCATGAATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
|||||
Db 172 GGATAAGACTGTCATGAATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 228
|||||

RESULT 7
AY026524
LOCUS
DEFINITION Mus musculus muscle-specific protein CSL (Csl) mRNA, complete cds.
ACCESSION AY026524
VERSION AY026524.1 GI:14575061
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C.,
Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F.,
Mohun,T. and Harvey,R.P.
TITLE The small muscle-specific protein Csl modifies cell shape and
promotes myocyte fusion in an insulin-like growth factor
1-dependent manner
JOURNAL J. Cell Biol. 153 (5), 985-998 (2001)
MEDLINE 21275706
PUBMED 11381084
REFERENCE
AUTHORS Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C.,
Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F.,
Mohun,T. and Harvey,R.P.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2001) Developmental Biology, Victor Chang Cardiac
Research Institute, 384 Victoria St, Darlinghurst, Sydney, New
South Wales 2010, Australia
FEATURES
Location/Qualifiers
source
1..787
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/chromosome="X"
gene
1..787
/genes="Csl"
CDS
206..463

```

```

/genes="Csl"
/notes="ortholog of Homo sapiens SMPX"
/codon_start=1
/product="muscle-specific protein CSL"
/protein_id="AAK07682.1"
/db_xref="GI:14575062"
/translation="MSKQPISNVRAIQANINIPMGAPRPGAGOPPRRKSTPTEBGA
PTTSEKPIPGMKFPGVFNLSIQNVKSELKFPVKGQ"
BASE COUNT 233 a 167 c 185 g 202 t
ORIGIN
Query Match 74.7%; Score 42.6; DB 10; Length 787;
Best Local Similarity 84.2%; Pred. No. 1.5e-05;
Matches 48; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGATAAGACTGTCATGAATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
|||||
Db 188 GGCTAAGACCTTGTGAATATGTCGAAGCAGCCAATTTCACGTCAGAGCCATCCAG 244
|||||

RESULT 8
AF364070
LOCUS
DEFINITION Mus musculus SMPX protein (Smpx) mRNA, complete cds.
ACCESSION AF364070
VERSION AF364070.1 GI:13940507
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
TITLE Identification, mapping, and genomic structure of a novel
X-chromosomal human gene (SMPX) encoding a small muscular protein
Hum. Genet. 105 (5), 506-512 (1999)
MEDLINE 20065879
PUBMED 10598820
REFERENCE
AUTHORS Patzak,D.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
Fleischmannstr. 42-44, D-17487 Greifswald, Germany
FEATURES
Location/Qualifiers
source
1..936
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..936
/gene="Smpx"
CDS
199..456
/genes="Smpx"
/codon_start=1
/product="SMPX protein"
/protein_id="AAK50398.1"
/db_xref="GI:13940508"
/translation="MSKQPISNVRAIQANINIPMGAPRPGAGOPPRRKSTPTEBGA
PTTSEKPIPGMKFPGVFNLSIQNVKSELKFPVKGQ"
polya_signal
506..911
/genes="Smpx"
/evidence=experimental
BASE COUNT 283 a 187 c 203 g 263 t
ORIGIN
Query Match 74.7%; Score 42.6; DB 10; Length 936;
Best Local Similarity 84.2%; Pred. No. 1.5e-05;
Matches 48; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGATAAGACTGTCATGAATATGTCGAACAGCCAGTTTCCATGTTAGAGCCATCCAG 57
|||||
Db 181 GGCTAAGACCTTGTGAATATGTCGAAGCAGCCAATTTCACGTCAGAGCCATCCAG 237
|||||

RESULT 9

```

MMU245772	943 bp	mRNA	linear	ROD 12-APR-2001
LOCUS	Mus musculus mRNA for stretch responsive muscle (X-chromosome)			
DEFINITION	protein (Srmx gene).			
ACCESSION	AJ245772			
VERSION	1			
KEYWORDS	Srmx gene; stretch responsive muscle (X-chromosome).			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 943)			
	Kemp,T.J., Sadusky,T.J., Simon,M., Brown,R., Eastwood,M., Sassoon,D.A. and Coulton,G.R.			
TITLE	Identification of a novel stretch-responsive skeletal muscle gene (Smpx)			
JOURNAL	Genomics 72 (3), 260-271 (2001)			
MEDLINE	21295047			
PubMed	11401441			
REFERENCE	2 (bases 1 to 943)			
AUTHORS	Kemp,T.J.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial College School Of Medicine, SAF Building, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM			
FEATURES	Location/Qualifiers			
source	1..943			
	/organism="Mus musculus"			
	/strain="C57BL/10"			
	/db xref="taxon:10090"			
	/tissue_type="skeletal muscle"			
gene	1..943			
	/gene="Srmx"			
5'UTR	1..214			
	/gene="Srmx"			
CDS	215..472			
	/gene="Srmx"			
	/codon_start=1			
	/product="stretch responsive muscle (X-chromosome)"			
	/protein_id="CAC08493.1"			
	/db xref="GI:10178963"			
	/translation="MSKQPISNVRAIQANINIPMGAFPGAGOPPRRKSTPETEEGA PTTSEKKKPIGKKKFGPVVNLSEIQNVKSELKFVPGSQ"			
3'UTR	473..943			
	/gene="Srmx"			
misc_feature	535..539			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	585..590			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	658..662			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	706..710			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	740..744			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	779..785			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
misc_feature	909..913			
	/gene="Srmx"			
	/function="mRNA destabilising motif"			
polyA_signal	923..928			
	/gene="Srmx"			
polyA_site	943			
	/gene="Srmx"			
BASE COUNT	278 a 192 c 210 g 262 t			1 others
ORIGIN				

```

Query Match          74.7%; Score 42.6; DB 10; Length 943;
Best Local Similarity 84.2%; Pred. No. 1.5e-05;
Matches 48; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAACACGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
|||||
Db 197 GGCTAAGACCTTGTGAATATGTCGAGCAGCCAAATTTCCAACTGTCAGAGCCATCCAG 253
|||||

RESULT 10
AF364071
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AF364071
Rattus norvegicus SMPX protein (Smpx) mRNA linear ROD 04-MAY-2001
complete cds.
AF364071.1 GI:13940509
Rattus norvegicus.
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 892)
Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
Identification, mapping, and genomic structure of a novel
X-chromosomal human gene (SMPX) encoding a small muscular protein
Hum. Genet. 105 (5), 506-512 (1999)
20065879
10598820
2 (bases 1 to 892)
Patzak,D.
Direct Submission
Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
Fleischmannstr. 42-44, D-17487 Greifswald, Germany
FEATURES
source
1..892
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
<1..892
/gene="Smpx"
190..447
/gene="Smpx"
/codon_start=1
/product="SMPX protein"
/protein_id="AAK50399.1"
/db_xref="GI:13940510"
/translations="MSGKPTSNVRSIQANINPMGAFRPGAGQPPRRKESKTPGTAEAGA"
PATPESEKKVPVGMKKFGPVPVNLSEIQNVKSELKYPVKGQG"
polya_signal
504..509
/gene="Smpx"
/evidence=experimental
polya_signal
862..867
/gene="Smpx"
/notes="alternate"
/evidence=experimental
BASE COUNT 271 a 183 c 193 g 245 t
ORIGIN

Query Match          69.1%; Score 39.4; DB 10; Length 892;
Best Local Similarity 80.7%; Pred. No. 0.00026;
Matches 46; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
|||||
Db 172 GTATAAGACCTCTTGAACATGTCGAAGCAGCCAAATTTCCAACTGTCAGATCCATTCAG 228
|||||

RESULT 11
AC124177
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC124177
Mus musculus chromosome UNK clone RP23-193C10, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC124177
AC124177.1 GI:21392586

```

KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 230579)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 230579)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (12-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 230579)
 AUTHORS McPherson,J.D. and Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu
 ----- Project Information -----
 Center project name: M_BA0193C10
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 228240 bases at least Q40
 Consensus quality: 228437 bases at least Q30
 Consensus quality: 228615 bases at least Q20
 Insert size: 220000; agarose-fp
 Insert size: 230079; sum-of-contigs
 Quality coverage: 13.39 in Q20 bases; agarose-fp
 Quality coverage: 11.47 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 9086: contig of 9086 bp in length
 * 9087 9186: gap of unknown length
 * 9187 20048: contig of 10862 bp in length
 * 20049 20148: gap of unknown length
 * 20149 40452: contig of 20304 bp in length
 * 40453 40552: gap of unknown length
 * 40553 79281: contig of 38729 bp in length
 * 79282 79381: gap of unknown length
 * 79382 137890: contig of 58509 bp in length
 * 137891 137990: gap of unknown length
 * 137991 230579: contig of 92589 bp in length.
 Location/Qualifiers
 1. 230579
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-193C10"
 1. 9086
 /note="assembly_name:Contig2"
 9187. 20048
 /note="assembly_name:Contig3"
 20149. 40452
 /note="assembly_name:Contig4"

misc_feature 40553. 79281
 /note="assembly_name:Contig5"
 79382. 137890
 /note="assembly_name:Contig6"
 misc_feature 137991. 230579
 /note="assembly_name:Contig7"
 BASE COUNT 66550 a 46384 c 47285 g 69857 t 503 others
 ORIGIN
 Query Match 52.3%; Score 29.8; DB 2; Length 230579;
 Best Local Similarity 70.2%; Pred. NO. 0.91;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 1 GGATAGACTGCATGATATGTCGAACACGACGTTCCCAATGTTAGACCCATCCAG 57
 Db 61167 GGATAATAATGCTTCTACATGTGGCAAGGAAGTTTATATTTTGGACCGCTTAG 61223
 RESULT 12
 AL589691/c
 LOCUS Human DNA sequence from clone RP11-80A23 on chromosome 6, complete
 DEFINITION sequence.
 ACCESSION AL589691
 VERSION AL589691.4 GI:14018306
 KEYWORDS HTG
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 44942)
 AUTHORS Blakey S.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 COMMENT On May 11, 2001 this sequence version replaced gi:1374045.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP11-80A23 is from the library RPC1-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-80A23 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP11-295F4 is at 44843 in this sequence.
 The true right end of clone RP11-560I21 is at 100 in this sequence.
 FEATURES
 Location/Qualifiers
 1. 44942
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"

Contact: humquery@anger.ac.uk
 ----- Project Information
 Center project name: BA435E4
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 146336 bases at least Q40
 Consensus quality: 155210 bases at least Q30
 Consensus quality: 160764 bases at least Q20
 Insert size: 185635; sum-of-contigs
 Insert size: 183829; 3.1% error; agarose-fp
 Quality coverage: 2.46x in Q20 bases; sum-of-contigs Quality
 coverage: 2.33x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 27 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 2836: contig of 2836 bp in length
 * 2837 2936: gap of 100 bp
 * 2937 9386: contig of 6450 bp in length
 * 9387 9486: gap of 100 bp
 * 9487 13859: contig of 4373 bp in length
 * 13860 13959: gap of 100 bp
 * 13960 21381: contig of 7422 bp in length
 * 21382 21481: gap of 100 bp
 * 21482 29095: contig of 7614 bp in length
 * 29096 29195: gap of 100 bp
 * 29196 37170: contig of 7975 bp in length
 * 37171 37270: gap of 100 bp
 * 37271 43364: contig of 6094 bp in length
 * 43365 43464: gap of 100 bp
 * 43465 48232: contig of 4768 bp in length
 * 48233 48332: contig of 4994 bp in length
 * 48333 53326: contig of 4994 bp in length
 * 53327 53426: gap of 100 bp
 * 53427 64045: contig of 10619 bp in length
 * 64046 64145: gap of 100 bp
 * 64146 73465: contig of 9320 bp in length
 * 73466 73565: gap of 100 bp
 * 73566 80267: contig of 6702 bp in length
 * 80268 80367: gap of 100 bp
 * 80368 83769: contig of 3402 bp in length
 * 83770 83869: gap of 100 bp
 * 83870 87023: contig of 3154 bp in length
 * 87024 87123: gap of 100 bp
 * 87124 94152: contig of 7029 bp in length
 * 94153 94252: gap of 100 bp
 * 94253 111698: contig of 17446 bp in length
 * 111699 111798: gap of 100 bp
 * 111799 114237: contig of 2439 bp in length
 * 114238 114337: gap of 100 bp
 * 114338 119111: contig of 4774 bp in length
 * 119112 119211: gap of 100 bp
 * 119212 125994: contig of 6783 bp in length
 * 125995 126094: gap of 100 bp
 * 126095 128771: contig of 2677 bp in length
 * 128772 128871: gap of 100 bp
 * 128872 132807: contig of 3936 bp in length
 * 132808 132907: gap of 100 bp
 * 132908 145229: contig of 12322 bp in length
 * 145230 145329: gap of 100 bp
 * 145330 148554: contig of 3225 bp in length
 * 148555 148654: gap of 100 bp
 * 148655 154822: contig of 6168 bp in length
 * 154823 154922: gap of 100 bp
 * 154923 157763: contig of 2841 bp in length
 * 157764 157863: gap of 100 bp

* 157864 163565: contig of 5702 bp in length
 * 163566 163665: gap of 100 bp
 * 163666 168235: contig of 4570 bp in length.
 FEATURES
 source
 1. 168235
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP11-435E4"
 /clone_lib="RPC1-11.2"
 1. 2836
 /note="assembly fragment:01401
 fragment chain:1"
 2937..9386
 /note="assembly fragment:00773
 fragment chain:1"
 9487..13859
 /note="assembly fragment:00477
 fragment chain:1"
 13960..21381
 /note="assembly fragment:01334
 fragment chain:1"
 21482..29095
 /note="assembly fragment:00150
 fragment chain:2"
 29196..37170
 /note="assembly fragment:00938
 fragment chain:2"
 37271..43364
 /note="assembly fragment:00292
 fragment chain:3"
 43465..48232
 /note="assembly fragment:00664
 fragment chain:3"
 48333..53326
 /note="assembly fragment:00712
 fragment chain:4"
 53427..64045
 /note="assembly fragment:01127
 fragment chain:4"
 64146..73465
 /note="assembly fragment:00897
 fragment chain:5"
 73566..80267
 /note="assembly fragment:00203
 fragment chain:5"
 80368..83769
 /note="assembly fragment:00039"
 83870..87023
 /note="assembly fragment:00121"
 87124..94152
 /note="assembly fragment:00170"
 94253..111698
 /note="assembly fragment:00240"
 111799..114237
 /note="assembly fragment:00315"
 114338..119111
 /note="assembly fragment:00376"
 119212..125994
 /note="assembly fragment:00398"
 126095..128771
 /note="assembly fragment:00812"
 128872..132807
 /note="assembly fragment:00943"
 132908..145229
 /note="assembly fragment:00977"
 145330..148554
 /note="assembly fragment:00998"
 148655..154822
 /note="assembly fragment:01019"
 154923..157763
 /note="assembly fragment:01169"
 157864..163565
 misc_feature

```

/note="assembly_fragment:01250"
163666..168235
/note="assembly_fragment:00613
clone_end:SP6
vector_side:right"
BASE COUNT 50800 a 29954 c 31524 g 53321 t 2636 others
ORIGIN
Query Match 47.0%; Score 26.8; DB 2; Length 168235;
Best Local Similarity 68.5%; Pred. No. 13;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4 TAAGACTGCATGAATATGCGAAGACGACGAGTTCCTCAATGTTAGAGCCATCCAG 57
|||||
Db 127688 TTAGCATGAATAATTTGCTAAAGACCCAGCTTGTATGTTCCAGCATCTAG 127741
|||||

RESULT 14
D26519/c
LOCUS
DEFINITION Mustela zibellina mitochondrial DNA for cytochrome b, partial cds.
ACCESSION D26519
VERSION D26519.1 GI:560116
KEYWORDS cytochrome b;
SOURCE Mustela zibellina muscle mitochondrion DNA.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
Martes.
REFERENCE
1. (bases 1 to 375)
Masuda, R. and Yoshida, M.C.
A molecular phylogeny of the family Mustelidae (Mammalia,
Carnivora), based on comparison of mitochondrial cytochrome b
nucleotide sequences
Zool. Sci. 11 (4), 605-612 (1994)

JOURNAL
MEDLINE 95038277
REFERENCE
2 (bases 1 to 375)
Masuda, R.
Direct Submission
Submitted (17-JAN-1994) Ryutichi Masuda, Hokkaido University,
Chromosome Research Unit, Faculty of Science, North 10, West 8,
Kita-ku, Sapporo, Hokkaido 060-0810, Japan
(E-mail:masudary@ees.hokudai.ac.jp, Tel:81-11-706-3541,
Fax:81-11-736-6304)
FEATURES
Location/Qualifiers
source
1..375
/organism="Mustela zibellina"
/organelle="mitochondrion"
/db_xref="taxon:36722"
/tissue_type="muscle"
1..>375
/codon_start=1
/transl_table=2
/product="cytochrome b"
/protein_id="BAA05525.1"
/db_xref="GI:577332"
/translations="MTNIRKTHPLAKIINNSFDLPAPSNISAWNFGSLIGLCILQ
ILTGLFLAMHYTSDTATPSSVTHICRDVNYGWIIRYMHANGASMEFFICLFLHVGRL
YIGSYNYPETWNIIGILLFAVMA"
BASE COUNT 103 a 107 c 57 g 108 t
ORIGIN
Query Match 46.7%; Score 26.6; DB 4; Length 375;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3 ATAAGACTGCATGAATATGCGAAGACGACGAGTTCCTCAATGTTAGAGCC 51
|||||
Db 148 ATAAACCTGTGAAGATCTGTAGGATCAGACAGATTCACAGAGGAGGCC 100
|||||

RESULT 15
L77957/c
LOCUS
DEFINITION Martes zibellina cytochrome b (cytb) gene, partial cds;
mitochondrial gene for mitochondrial product.
ACCESSION L77957
VERSION L77957.1 GI:6670761
KEYWORDS cytochrome b;
SOURCE Martes zibellina.
ORGANISM
Mitochondrion Martes zibellina
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
Martes.
REFERENCE
1. (bases 1 to 401)
Carr, S.M. and Hicks, S.A.
Are there two species of pine marten in North America? Genetic and
evolutionary relationships within Martes (Eds.);
(in) Proulx, G., Goddard, R. and Bryant, H. (Eds.);
MARTES: TAXONOMY, ECOLOGY, TECHNIQUES, AND MANAGEMENT: 15-28;
Provincial Museum of Alberta, Edmonton (1997)
2 (bases 1 to 401)
Carr, S.M.
Direct Submission
Submitted (01-APR-1997) Department of Biology, Memorial University
of Newfoundland, St. John's, Newfoundland A1B 3X9, Canada
GSDS:S:78288
Location/Qualifiers
source
1..401
/organism="Martes zibellina"
/organelle="mitochondrion"
/db_xref="taxon:36722"
/dev_stage="adult"
1..>401
/gene="cytb"
1..>401
/gene="cytb"
/note="putative"
/codon_start=1
/transl_table=2
/product="cytochrome b"
/protein_id="AAF23046.1"
/db_xref="GI:6670762"
/translations="MTNIRKTHPLAKIINNSFDLPAPSNISAWNFGSLIGLCILQ
ILTGLFLAMHYTSDTATPSSVTHICRDVNYGWIIRYMHANGASMEFFICLFLHVGRL
YIGSYNYPETWNIIGILLFAVMA"
exon
1..>401
/gene="cytb"
/note="putative"
BASE COUNT 110 a 113 c 61 g 117 t
ORIGIN
Query Match 46.7%; Score 26.6; DB 4; Length 401;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3 ATAAGACTGCATGAATATGCGAAGACGACGAGTTCCTCAATGTTAGAGCC 51
|||||
Db 148 ATAAACCTGTGAAGATCTGTAGGATCAGACAGATTCACAGAGGAGGCC 100
|||||

Search completed: April 15, 2003, 20:55:10
Job time : 248.143 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:52:31 ; Search time 118.221 Seconds
(without alignments)
7808.593 Million cell updates/sec

Title: US-09-647-019-7

Perfect score: 57

Sequence: 1 ggataagactgatgaatat.....ccaatgttagccatccag 57

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpi:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	206	14	R58129 F9260 Petal
2	57	100.0	214	9	AA092554 115591.se
3	57	100.0	225	9	AA249531 JJ5022.se
4	57	100.0	225	14	N56276 JJ5022P Hum
5	57	100.0	284	9	AA094015 cl1599.se
6	57	100.0	360	9	AI355905 3055-10R-

7	57	100.0	501	9	AA211521	AA211521 zn55b01.r
8	57	100.0	719	14	BM697544	BM697544 UI-B-DX0-
9	57	100.0	756	12	BF791178	BF791178 602251278
10	57	100.0	793	12	BE671987	BE671987 602152408
11	57	100.0	826	12	BF128275	BF128275 601650456
12	57	100.0	855	12	BF672902	BF672902 602152759
13	57	100.0	867	12	BF790636	BF790636 602250354
14	57	100.0	871	12	BF693124	BF693124 602080151
15	57	100.0	879	12	BF575112	BF575112 602134792
16	57	100.0	906	12	BF672126	BF672126 602152580
17	57	100.0	909	12	BF790243	BF790243 602249777
18	50.6	88.8	261	9	AU083330	AU083330 AU083330
19	49	86.0	517	12	BG224245	BG224245 1M0028F05
20	49	86.0	521	12	BF261282	BF261282 1M0004B05
21	49	86.0	528	12	BF073866	BF073866 220906 MA
22	49	86.0	614	12	BG223758	BG223758 1M00018G0
23	48	83.2	466	12	BG358790	BG358790 BOWMS1-01
24	47.4	83.2	278	12	BF429539	BF429539 1075 MARC
25	47	82.5	412	12	BG223652	BG223652 1M00010F0
26	46	80.7	320	14	W07478	W07478 za96c10.r1
27	46	80.7	394	9	AA389647	AA389647 M104 Feta
28	45.8	80.4	544	10	BE014221	BE014221 125832 MA
29	45	78.9	280	9	AA248485	AA248485 chn0287.s
30	45	78.9	319	14	N87511	N87511 LL1191F Hum
31	43.4	76.1	185	9	AA247862	AA247862 J3408.seq
32	42.6	74.7	265	14	W14689	W14689 mb34a08.r1
33	42.6	74.7	330	9	AA763276	AA763276 vv89a04.r
34	42.6	74.7	364	14	W97451	W97451 mf97C02.r1
35	42.6	74.7	368	14	W29186	W29186 mc22f02.r1
36	42.6	74.7	389	9	AI098485	AI098485 uc06c10.r
37	42.6	74.7	416	9	AI153970	AI153970 ud49h02.r
38	42.6	74.7	436	14	W36988	W36988 mb65b11.r1
39	42.6	74.7	448	14	W13738	W13738 mb32a12.r1
40	42.6	74.7	465	14	W18646	W18646 mb98a05.r1
41	42.6	74.7	466	14	W18392	W18392 mb98e01.r1
42	42.6	74.7	468	14	W29828	W29828 mc07b11.r1
43	42.6	74.7	482	9	AA434782	AA434782 ve23c01.r
44	42.6	74.7	486	9	AI035961	AI035961 ub50b12.r
45	42.6	74.7	490	9	AA060214	AA060214 mj65h06.r

ALIGNMENTS

RESULT 1

R58129 F9260 Fetal heart Homo sapiens cDNA clone F9260 5' end, mRNA
sequence.
R58129
VERSION R58129.1 GI:828187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 206)
AUTHORS Hwang, D.M., Fung, Y.W., Wang, R.X., Laurensen, C.M., Ng, S.H., Lam, W.Y., Tsui, K.W., Fung, K.P., Waye, M., Lee, C.Y. and Liaw, C.C.
TITLE Analysis of expressed sequence tags from a fetal human heart cDNA library
JOURNAL Genomics 30 (2), 293-298 (1995)
MEDLINE 96163883
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 617750995
Email: cliw@rics.bwh.harvard.edu
Seq primer: GTGGCGACGACTCTGTGAGCC.
Location/Qualifiers
source 1. .206

RESULT 3	
AA249531	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				
41				
42				
43				
44				
45				
46				
47				
48				
49				
50				
51				
52				
53				
54				
55				
56				
57				
58				
59				
60				
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				
71				
72				
73				
74				
75				
76				
77				
78				
79				
80				
81				
82				
83				
84				
85				
86				
87				
88				
89				
90				
91				
92				
93				
94				
95				
96				
97				
98				
99				
100				

FEATURES source

/db_xref="taxon:9606"
 /clone_lib="Human fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-Oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."

BASE COUNT
ORIGIN

Query Mat
Best Loca
Matches

SOURCE	ORGANISM
REFERENCE	AUTHORS
TITLE	JOURNAL
COMMENT	

Contact: Llew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA

Tel: 6177328915
Fax: 6179750995

Email: cliw@rics.bwh.harvard.edu
Seq primer: TCCAAAGAAATTCGCACGAG.

Location/Qualifiers

FEATURES

source

1. .225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="JJ5022"
/lab_host="E. coli XLI-Blue"
/note="Vector: Lambda ZAP Express"
XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 63 a 49 c 57 g 56 t
ORIGIN
Query Match 100.0%; Score 57; DB 14; Length 225;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57
|||||
Db 100 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 156
|||||

RESULT 5

AA094015 284 bp mRNA linear EST 25-OCT-1996
LOCUS cl1599.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens
DEFINITION cDNA 5', mRNA sequence.

ACCESSION AA094015

VERSION AA094015.1 GI:1639600

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 284)

REFERENCE Liew,C.C.

TITLE cDNAs from fetal heart (1996)

JOURNAL Unpublished (1996)

COMMENT Contact: Liew CC

Brigham and Women's Hospital

Harvard Medical School

75 Francis St. Boston, MA 02115, USA

Tel: 6177328915

Fax: 6179750995

Email: cliw@rics.bwh.harvard.edu

PCR Primers

FORWARD: 5' GCCAAGCTCGAATTAACCTCTCACTAAAGG 3'

BACKWARD: 5' CCAGTAATGTAACGACCTCACTATAGGGG 3'

Seq primer: 5' GAAATTAACCTCTCACTAAAGG 3'.

Location/Qualifiers

1. .284

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Human fetal heart, Lambda ZAP Express"

/lab_host="E. coli XLI-Blue"

/notes="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 78 a 74 c 70 g 61 t

ORIGIN

Query Match 100.0%; Score 57; DB 9; Length 284;

Best Local Similarity 100.0%; Pred. No. 3.4e-10;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57

|||||

Db 152 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 208

|||||

Matches

57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57

|||||

Db 87 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 143

|||||

RESULT 6

AI355905 360 bp mRNA linear EST 31-DEC-1999
LOCUS 3055-10R-8H11-2 Human heart cDNA (CCLee) Homo sapiens cDNA clone
DEFINITION xh811.3', mRNA sequence.

ACCESSION AI355905

VERSION AI355905.1 GI:6649247

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 360)

REFERENCE Zhuchenko O., Patzak D. and Wehnert M.

AUTHORS ESTs of clone xh811 (human heart)

JOURNAL Unpublished (1999)

COMMENT Other ESTs: 8H11-R1-C8-OZ, 8H11-F1-C7-OZ, 3055-10U-8H11-2

Contact: Patzak D., submitter; (Wehnert M.; supervisor)

Department of Molecular Human Genetics

Institut for Human Genetics

Fleischmannstr. 42/44, D-17487 Greifswald, Germany

Tel: +49 3834 8653-78 (-74)

Fax: +49 3834 8653-93

Email: patzak@uni-greifswald.de (mwehnert@uni-greifswald.de)

Single path sequence (manual checked and edited) with a vector

primer of clone xh811. Sequence starts with base 1 following the

determined vector sequence (ACCGGTGCGCGCT, small letters:

NotI-site). ESTs 3055-10R-8H11-2, 3055-10U-8H11-2 (with vector

primers) and 8H11-R1-C8-OZ, 8H11-U1-C7-OZ (with gene specific

primers) provide a contig of the whole insert of clone xh811.

Insert Length: 850 Std Error: 10.00

Plate: 8 row: H column: 11

Seq primer: 3055-10R (like M13 reverse)

High quality sequence stop: 360.

Location/Qualifiers

1. .360

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="xh811"

/clone_lib="Human heart cDNA (CCLee)"

/sex="male"

/tissue_type="heart muscle"

/dev_stage="adult"

/lab_host="E. coli"

/note="Organ: heart; Vector: Lambda ZAP II, pKSII-;

Site 1: NotI; Site 2: Human heart cDNA library was

constructed in Lambda ZAP II vectors using NotI linkers.

Clones from the primary cDNA library were deposited into

36-well trays for storage and retrieval. The isolation of

chromosome-specific genes by reciprocal probing of arrayed

cDNA and cosmid libraries' (Human Molecular Genetics, 1995

, Vol. 4, No 8; pp1373-1380) provided cDNA-Clones as

plasmids (vector pKSII-, E. coli)."

BASE COUNT 109 a 92 c 88 g 71 t

ORIGIN

Query Match 100.0%; Score 57; DB 9; Length 360;

Best Local Similarity 100.0%; Pred. No. 3.6e-10;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 57

|||||

Db 152 GGATAAGACTGCATGATATGTCGAAACAGCCAGTTTCCAAATGTTAGAGCCATCCAG 208

|||||

RESULT 7
AA211521
LOCUS
DEFINITION
zn55D01.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
IMAGE:562057 5', mRNA sequence.

ACCESSION
AA211521
VERSION
AA211521.1 GI:1810175
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 501)

AUTHORS
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996).

TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 443.

FEATURES

source

1..501
/organism="Homo sapiens"
/db_xref="GDB:4595347"
/db_xref="taxon:9606"
/clone="IMAGE:562057"
/clone_lib="Stratagene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Skeletal muscle from patient with
malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG
3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 167 a 103 c 116 g 115 t

ORIGIN

Query Match 100.0%; Score 57; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAGACTGCATGATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCATCCAG 57
|||||
Db 86 GGATAGACTGCATGATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCATCCAG 142
|||||

RESULT 8
BM697544
LOCUS
DEFINITION
UI-E-DX0-agn-i-12-0-UI.r1 UI-E-DX0 Homo sapiens cDNA clone
IMAGE:562057 5', mRNA sequence.

ACCESSION
BM697544
VERSION
BM697544.1 GI:19010802
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 719)

AUTHORS

TITLE

JOURNAL
MEDLINE
COMMENT

Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing By: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
source

1..719
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-DX0-agn-i-12-0-UI"
/clone_lib="UI-E-DX0"
/tissue_type="fetal eyes"
/dev_stage="fetal"
/lab_host="DH10B (life technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-DX0 is a cDNA library containing the following
tissue(s): fetal eyes. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is AGAATCAAGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 226 a 142 c 163 g 186 t

ORIGIN

Query Match 100.0%; Score 57; DB 14; Length 719;
Best Local Similarity 100.0%; Pred. No. 4.2e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAGACTGCATGATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCATCCAG 57
|||||
Db 128 GGATAGACTGCATGATATGTCGAACAGCCAGTTCCCAATGTTAGAGCCATCCAG 184
|||||

RESULT 9
BF791178
LOCUS
DEFINITION
602251278F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4338594 5',
mRNA sequence.

ACCESSION
BF791178
VERSION
BF791178.1 GI:12096232
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 756)

AUTHORS
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1214 row: 1 column: 19
 High quality sequence stop: 619.
 Location/Qualifiers

FEATURES

source
 1. .756
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4338594"
 /clone_lib="NIH MGC 81"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site 1: SfiI (ggccgctcgcc); Site 2: SfiI
 (ggccattatggcc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 243 a 158 c 170 g 185 t
 ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 756;
 Best Local Similarity 100.0%; Pred. No. 4.2e-10;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAGACTGCATGAATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57
 |||||
 Db 134 GGATAGACTGCATGAATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 190
 |||||

RESULT 10
 BF671987
 LOCUS
 DEFINITION 793 bp mRNA linear EST 21-DEC-2000
 602152408F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293721 5',
 mRNA sequence.
 ACCESSION BF671987
 VERSION BF671987.1 GI:11945882
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 793)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM1141 row: 0 column: 02
 High quality sequence stop: 665.
 Location/Qualifiers

FEATURES

source
 1. .793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4293721"
 /clone_lib="NIH MGC 81"
 /lab_host="DH10B (TI phage-resistant)"

/note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site 1: SfiI (ggccgctcgcc); Site 2: SfiI
 (ggccattatggcc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 251 a 164 c 175 g 203 t
 ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 793;
 Best Local Similarity 100.0%; Pred. No. 4.3e-10;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATAGACTGCATGAATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 57
 |||||
 Db 135 GGATAGACTGCATGAATATGTCGAACAGCCAGTTCCTCAATGTTAGAGCCATCCAG 191
 |||||

RESULT 11
 BF126275
 LOCUS
 DEFINITION 836 bp mRNA linear EST 24-OCT-2000
 601650456F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934333 5',
 mRNA sequence.
 ACCESSION BF126275
 VERSION BF126275.1 GI:10965233
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 836)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM774 row: p column: 14
 High quality sequence stop: 651.
 Location/Qualifiers

source
 1. .826
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3934333"
 /clone_lib="NIH MGC 76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1:
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCCGACATG-dT(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH MGC Library."
 BASE COUNT 280 a 156 c 188 g 202 t
 ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 826;
 Best Local Similarity 100.0%; Pred. No. 4.3e-10;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 57
|||||
Db 82 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 138

RESULT 12
BF672902 855 bp mRNA linear EST 21-DEC-2000
LOCUS 602152759P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293878 5',
DEFINITION mRNA sequence.
ACCESSION BF672902
VERSION BF672902.1 GI:11946797
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 855)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1142 row: e column: 15
High quality sequence stop: 687.
FEATURES
source
1. .855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4293878"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
(ggccaattggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
BASE COUNT 270 a 159 c 210 g 216 t
ORIGIN
Query Match 100.0%; Score 57; DB 12; Length 855;
Best Local Similarity -100.0%; Pred. No. 4.3e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 57
|||||
Db 101 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 157

RESULT 13
BF790636 867 bp mRNA linear EST 12-JAN-2001
LOCUS 602250354F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4337821 5',
DEFINITION mRNA sequence.
ACCESSION BF790636
VERSION BF790636.1 GI:12095690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 867)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1056 row: f column: 13

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 867)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1212 row: 1 column: 14
High quality sequence stop: 514.

FEATURES
source

1. .867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4337821"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site 1: SfiI (ggcgctcgcc); Site 2: SfiI
(ggccaattggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

BASE COUNT 248 a 165 c 244 g 210 t
ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 867;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 57
|||||
Db 224 GGATAAGACTGCATGAATATGTCGAAACAGCCAGTTTCCAATGTTAGAGCCATCCAG 280

RESULT 14
BF693124

LOCUS 602080151F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4244364 5',
DEFINITION mRNA sequence.
ACCESSION BF693124
VERSION BF693124.1 GI:11978532
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 871)
NIH-MGC http://mgc.nci.nih.gov/.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1056 row: f column: 13

High quality sequence stop: 575.

FEATURES

source

1..871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:424364"
/lab_host="NIH MGC 81"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 280 a 157 c 218 g 216 t

ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 871;
Best Local Similarity 100.0%; Pred. No. 4.4e-10;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGATATGCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57
|||||
Db 134 GGATAAGACTGCATGATATGCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 190

RESULT 15

BF575112

LOCUS

DEFINITION 602134792F1 NIH_MGC_81 879 bp mRNA linear EST 12-DEC-2000
mRNA sequence.

ACCESSION BF575112

VERSION BF575112.1 GI:11648824

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-f@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.

CDNA Library Preparation: CLONETECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Invitrogen Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM131 row: h column: 24

High quality sequence stop: 595.

FEATURES

source

1..879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4289735"
/clone_lib="NIH MGC 81"
/lab_host="NIH MGC 81"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained

inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 284 a 147 c 239 g 208 t 1 others

ORIGIN

Query Match 100.0%; Score 57; DB 12; Length 879;

Best Local Similarity 100.0%; Pred. No. 4.4e-10;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGATAAGACTGCATGATATGCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 57

|||||

Db 133 GGATAAGACTGCATGATATGCGAAACAGCCAGTTTCCCAATGTTAGAGCCATCCAG 189

|||||

Search completed: April 15, 2003, 22:02:31

Job time : 119.221 secs

Vol. 4, No 8; pp1373-1380) provided cDNA-Clones as plasmids (vector pksf1-, E. coli)."

BASE COUNT 109 a 92 c 88 g 71 t
 ORIGIN
 Query Match 100.0%; Score 87; DB 9; Length 360;
 Best Local Similarity 100.0%; Pred. No. 6.3e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCCTGAGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCGAGAAGA 60
 DB 209 GCAATATCAATATTCCTGAGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCGAGAAGA 268

QY 61 AAGAATGTAATCTGCTGAAGTGGAGGAG 87
 DB 269 AAGAATGTAATCTGCTGAAGTGGAGGAG 295

RESULT 4
 AA214155
 LOCUS
 DEFINITION zn58f10.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
 ACCESSION AA214155
 VERSION 'AA214155.1 GI:1812792
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 480)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, P., Thierry-Mieg, J., Trevaaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LBNL; contact the
 IMAGE Consortium (info@image.lbnl.gov) for further information.
 Insert Length: 852 Std Error: 0.00
 Seq primer: -28m3 rev1 ET from Amersham
 High quality sequence stop: 191.

FEATURES
 source
 1..480
 /organism="Homo sapiens"
 /db_xref="GDB:4595701"
 /db_xref="taxon:9606"
 /clone="IMAGE:562411"
 /clone_lib="Stratagene muscle 937209"
 /tissue_type="muscle"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skeletal muscle; Vector: pBluescript SK-;
 Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.
 Primer: Oligo dT. Skeletal muscle from patient with
 malignant hyperthermia. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG
 3' -3' adaptor sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3' "

BASE COUNT 146 a 96 c 113 g 112 t 13 others
 ORIGIN

Query Match 100.0%; Score 87; DB 9; Length 480;
 Best Local Similarity 100.0%; Pred. No. 7e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCCTGAGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCGAGAAGA 60
 DB 156 GCAATATCAATATTCCTGAGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCGAGAAGA 215

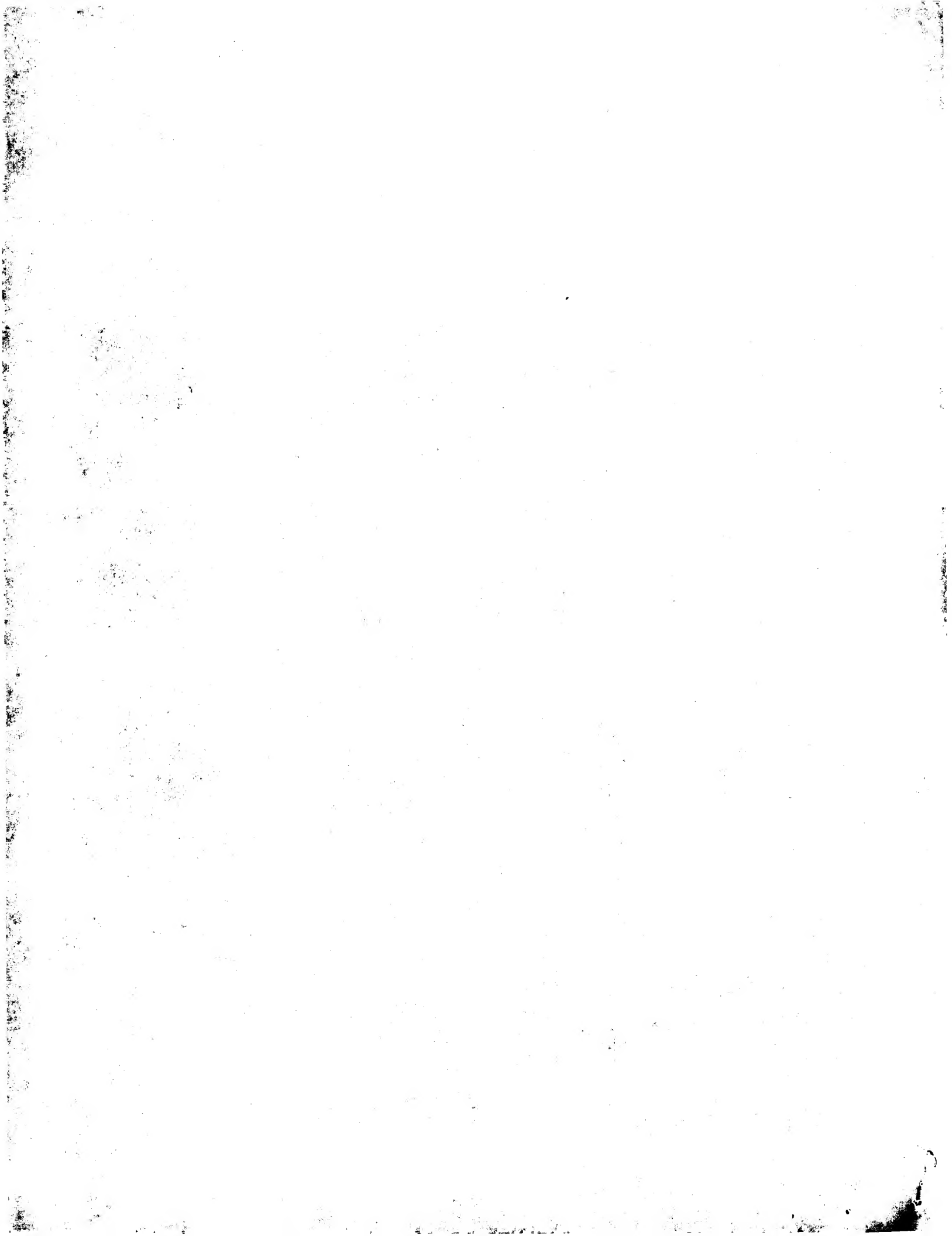
QY 61 AAGAATGTAATCTGCTGAAGTGGAGGAG 87
 DB 216 AAGAATGTAATCTGCTGAAGTGGAGGAG 242

RESULT 5
 BM697544
 LOCUS
 DEFINITION UI-E-DX0-agn-i-12-0-UI.r1 UI-E-DX0 Homo sapiens cDNA clone
 ACCESSION BM697544
 VERSION UI-E-DX0-agn-i-12-0-UI 5', mRNA sequence.
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 719)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Reverse.

FEATURES
 source
 1..719
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-DX0-agn-i-12-0-UI"
 /clone_lib="UI-E-DX0"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DX0 is a cDNA library containing the following
 tissue(s): fetal eyes. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is AGAATCAAGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT 226 a 142 c 163 g 186 t 2 others
 ORIGIN

Query Match 100.0%; Score 87; DB 14; Length 719;
 Best Local Similarity 100.0%; Pred. No. 7.9e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Wed Apr 16 05:43:54 2003

us-09-64

Qy	1	GCAATATCAATATTC	60
Db	185	GCAATATCAATATTC	244
Qy	61	AAAGANTGTACTCT	87
Db	245	AAAGANTGTACTCT	271

RESULT 6

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:52:31 ; Search time 180.443 Seconds
(without alignments)
7808.593 Million cell updates/sec

Title: US-09-647-019-8
Perfect score: 87
Sequence: 1 gcaaatcaatattccaat.....gtactctgaatggaggag 87

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1:	em_estba:*
2:	em_esthum:*
3:	em_estlin:*
4:	em_estnu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hctc:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hct:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_ges:*
18:	em_ges_hum:*
19:	em_ges_inv:*
20:	em_ges_pln:*
21:	em_ges_vrt:*
22:	em_ges_fun:*
23:	em_ges_nam:*
24:	em_ges_mus:*
25:	em_ges_other:*
26:	em_ges_pro:*
27:	em_ges_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	206	14 R58129	R58129 F9260 Fetal
2	87	100.0	320	14 W07478	W07478 za96C10.r1
3	87	100.0	360	9 A1355905	A1355905 30S5-10R-
4	87	100.0	480	9 AA214155	AA214155 zn59f10.r
5	87	100.0	719	14 BM697544	BM697544 UI-E-DX0-
6	87	100.0	756	12 BF791178	BF791178 602251278

7	87	100.0	793	12	BF671987	BF671987 602152408
8	87	100.0	826	12	BF126275	BF126275 601650456
9	87	100.0	855	12	BF672902	BF672902 602152759
10	87	100.0	879	12	BF575112	BF575112 602134792
11	87	100.0	906	12	BF672126	BF672126 602152580
12	87	100.0	909	12	BF790243	BF790243 602249777
13	79	90.8	261	9	AU083330	AU083330 AU083330
14	76	87.4	867	12	BF790636	BF790636 602250354
15	75	86.2	394	9	AA389647	AA389647 M104 Feta
16	75	86.2	501	9	AA211521	AA211521 zn55b01.r
17	74.2	85.3	412	12	BG223652	BG223652 1M00010F0
18	74.2	85.3	466	12	BG358790	BG358790 BOWMS1-01
19	74.2	85.3	517	12	BG224245	BG224245 1M0028F05
20	74.2	85.3	521	12	BF261282	BF261282 1M0004B05
21	74.2	85.3	528	12	BF073866	BF073866 220906 MA
22	74.2	85.3	582	12	BG224246	BG224246 1M0030E01
23	74.2	85.3	614	12	BG223758	BG223758 1M00018G0
24	73.4	84.4	683	12	BE856875	BE856875 7F70C03.X
25	72.6	83.4	319	14	N87511	N87511 LUL131F Hum
26	72.6	83.4	544	10	BE044221	BE044221 125932 MA
27	71	81.6	278	12	BF429539	BF429539 1075 MARC
28	71	81.6	426	10	BE013988	BE013988 125544 MA
29	69	79.3	871	12	BF693124	BF693124 602080151
30	66.2	76.1	364	14	W97451	W97451 mf97c02.r1
31	66.2	76.1	368	14	W29186	W29186 mc22f02.r1
32	66.2	76.1	389	9	A1098485	A1098485 uc06c10.r
33	66.2	76.1	416	9	A1153970	A1153970 ud49h02.r
34	66.2	76.1	436	14	W36988	W36988 mb65b11.r1
35	66.2	76.1	448	14	W13738	W13738 mb32a12.r1
36	66.2	76.1	465	14	W18646	W18646 mb88a05.r1
37	66.2	76.1	466	14	W18392	W18392 mb88e01.r1
38	66.2	76.1	468	14	W29828	W29828 mc07b11.r1
39	66.2	76.1	482	9	AA434782	AA434782 ve23c01.r
40	66.2	76.1	486	9	A1035961	A1035961 mb50b12.r
41	66.2	76.1	504	9	AA033164	AA033164 ub137c11.r
42	66.2	76.1	505	9	A1787904	A1787904 ul14f07.y
43	66.2	76.1	540	11	AK009857	AK009857 Mus muscu
44	66.2	76.1	613	14	BQ554133	BQ554133 H4026C08-
45	66.2	76.1	617	12	BG794218	BG794218 UTSW_SM1G

ALIGNMENTS

RESULT 1
R58129
LOCUS F9260 Fetal heart Homo sapiens cDNA clone F9260 5' end, mRNA
DEFINITION 206 bp mRNA linear EST 02-MAY-1996
ACCESSION R58129
VERSION R58129.1 GI:828187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 206)
AUTHORS Huang,D.M., Fung,Y.W., Wang,R.X., Laurensen,C.M., Ng,S.H., Lam
W.Y., Teui,K.W., Fung,K.P., Waye,M., Lee,C.Y. and Liew,C.C.
TITLE Analysis of expressed sequence tags from a fetal human heart cDNA
library
JOURNAL Genomics 30 (2), 293-298 (1995)
MEDLINE 96163883
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
Seq primer: GGTGGGACGACTCTCGAGCC.
Location/Qualifiers 1. .206

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Fetal heart"
/lab_host="F. coli Y1090"
/note="Vector: Lambda gt22; Site 1: NotI; Site 2: SalI;
mRNA was purified from human fetal hearts (10-12 weeks).
cDNA was constructed using a NotI-Oligo dr adaptor-primer.
SalI adaptors were ligated, followed by digestion with
NotI, for direction cloning into predigested lambda gt22.
Method is described in J. Mol. Cell. Cardiol. (1994) 26,
1329-1333."

BASE COUNT 64 a 46 c 51 g 45 t
ORIGIN
Query Match 100.0%; Score 87; DB 14; Length 206;
Best Local Similarity 100.0%; Pred. No. 5.3e-17; Indels 0; Gaps 0;
Matches 87; Conservative 0; Mismatches 0;
QY 1 GCAATATCAATATTCATGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 60
Db 118 GCAATATCAATATTCATGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 177
QY 61 AAGAATGTACTCTGAAGTCGAGGAG 87
Db 178 AAGAATGTACTCTGAAGTCGAGGAG 204

RESULT 2
LOCUS W07478 320 bp mRNA linear EST 25-APR-1996
DEFINITION za96c10.r1 Soares_fetal_lung NBHL19W Homo sapiens cDNA clone
IMAGE:300402 5', mRNA sequence.
W07478
ACCESSION W07478.1 GI:1281479
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 320)
AUTHORS Hillier L., Clark N., Dubuque T., Elliston K., Hawkins M., Holman M., Hultman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J., Rifkin L., Rohlfing I., Soares M., Tan F., Trevaskis E., Waterston R., Williamson A., Wohlmann P., and Wilson R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 279.

FEATURES
source
1. .320
/organism="Homo sapiens"
/db_xref="GDB:1245326"
/db_xref="taxon:9606"
/clone_lib="Soares_fetal_lung_NBHL19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTCGAGGAGCGGCAATTTTTTTTTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Patima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBHL19W."

BASE COUNT 92 a 83 c 77 g 67 t 1 others
ORIGIN
Query Match 100.0%; Score 87; DB 14; Length 320;
Best Local Similarity 100.0%; Pred. No. 6.1e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAATATCAATATTCATGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 60
Db 212 GCAATATCAATATTCATGGAGCCTTTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 271
QY 61 AAGAATGTACTCTGAAGTCGAGGAG 87
Db 272 AAGAATGTACTCTGAAGTCGAGGAG 298

RESULT 3
LOCUS AI355905 360 bp mRNA linear EST 31-DEC-1999
DEFINITION 3055-10R-8H11-2 Human heart cDNA (CCLee) Homo sapiens cDNA clone
xh8H11.3', mRNA sequence.
AI355905
ACCESSION AI355905
VERSION AI355905.1 GI:6649247
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 360)
AUTHORS Zhuchenko O., Patzak D. and Wehnert M.
TITLE ESTs of clone xh8H11 (human heart)
JOURNAL Unpublished (1999)
COMMENT Other ESTs: 8H11-R1-C8-OZ, 8H11-F1-C7-OZ, 3055-10U-8H11-2
Contact: Patzak D., submitter: (Wehnert M.; supervisor)
Department of Molecular Human Genetics
Institut für Human Genetics
Fleischmannstr. 42/44, D-17487 Greifswald, Germany
Tel: +49 3834 8653-78(-74)
Fax: +49 3834 8653-93
Email: patzak@uni-greifswald.de (mwehnert@uni-greifswald.de)

Single path sequence (manual checked and edited) with a vector
primer of clone xh8H11. Sequence starts with base 1 following the
determined vector sequence (ACCGGGTGGCGCGCT, small letters:
NotI-site). ESTs 3055-10R-8H11-2, 3055-10U-8H11-2 (with vector
primers) and 8H11-R1-C8-OZ, 8H11-U1-C7-OZ (with gene specific
primers) provide a contig of the whole insert of clone xh8H11.
Insert Length: 850 Std Error: 10.00
Plate: 8 row: H column: 11
Seq primer: 3055-10R (like M13 reverse)
High quality sequence stop: 360.
Location/Qualifiers
1. .360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="xh8H11"
/clone_lib="Human heart cDNA (CCLee)"
/sex="male"
/tissue_type="heart muscle"
/dev_stage="adult"
/lab_host="F. coli"

FEATURES
source
1. .360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="xh8H11"
/clone_lib="Human heart cDNA (CCLee)"
/sex="male"
/tissue_type="heart muscle"
/dev_stage="adult"
/lab_host="F. coli"
/note="Organ: heart; Vector: Lambda ZAP II, pKSI1-;
Site 1: NotI; Site 2: NotI; Human heart cDNA library was
constructed in Lambda ZAP II vectors using NotI linkers.
Clones from the primary cDNA library were deposited into
96-well trays for storage and retrieval. The 'isolation of
chromosome-specific genes by reciprocal probing of arrayed
cDNA and cosmid libraries' (Human Molecular Genetics, 1995

BASE COUNT 109 a 92 c 88 g 71 t
 ORIGIN

Query Match 100.0%; Score 87; DB 9; Length 360;
 Best Local Similarity 100.0%; Pred. No. 6.3e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGAGCCTTTCCGCCAGGAGCAGGTCAACCCCAAGA 60
 |||
 Db 209 GCAATATCAATATTCATGGAGCCTTTCCGCCAGGAGCAGGTCAACCCCAAGA 268
 |||

QY 61 AAGAATGTACTCTGAAGTGGAGGAG 87
 |||
 Db 269 AAGAATGTACTCTGAAGTGGAGGAG 295
 |||

RESULT 4
 AA214155 480 bp mRNA linear EST 01-AUG-1997
 LOCUS zn58f10.r1 Stratagene muscle 937209 Homo sapiens cDNA clone
 DEFINITION IMAGE:562411 5', mRNA sequence.

ACCESSION AA214155
 VERSION AA214155.1 GI:1812792
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 480)

REFERENCE Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
 B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Warra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478

COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 852 Std Error: 0.00
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 191.

FEATURES
 source
 1..480
 /organism="Homo sapiens"
 /db_xref="GB:4595701"
 /db_xref="taxon:9606"
 /clone="IMAGE:562411"
 /clone_lib="Stratagene muscle 937209"
 /tissue_type="muscle"
 /dev_stage="adult"
 /lab_host="SOLR (kanamycin resistant)"

/note="Organ: skeletal muscle; Vector: pBluescript SK-;
 Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.
 Primer: Oligo dT. Skeletal muscle from patient with
 malignant hyperthermia. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGCGACGAG
 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3"

BASE COUNT 146 a 96 c 113 g 112 t 13 others

Query Match 100.0%; Score 87; DB 9; Length 480;
 Best Local Similarity 100.0%; Pred. No. 7e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGAGCCTTTCCGCCAGGAGCAGGTCAACCCCAAGA 60
 |||
 Db 156 GCAATATCAATATTCATGGAGCCTTTCCGCCAGGAGCAGGTCAACCCCAAGA 215
 |||

QY 61 AAGAATGTACTCTGAAGTGGAGGAG 87
 |||
 Db 216 AAGAATGTACTCTGAAGTGGAGGAG 242
 |||

RESULT 5
 BM697544 719 bp mRNA linear EST 28-FEB-2002
 LOCUS UI-E-DX0-agn-i-12-0-UI.r1 UI-E-DX0 Homo sapiens cDNA clone
 DEFINITION UI-E-DX0-agn-i-12-0-UI 5', mRNA sequence.

ACCESSION BM697544
 VERSION BM697544.1 GI:19010802
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 719)

REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477

COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msquares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com).
 Seq primer: M13 Reverse.

FEATURES
 source
 1..719
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-DX0-agn-i-12-0-UI"
 /clone_lib="UI-E-DX0"
 /tissue_type="fetal eyes"
 /dev_stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DX0 is a cDNA library containing the following
 tissue(s): fetal eyes. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT7T3-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is AGAATCAAGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT 226 a 142 c 163 g 186 t 2 others

Query Match 100.0%; Score 87; DB 14; Length 719;
 Best Local Similarity 100.0%; Pred. No. 7.9e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 60
 |||||
 Db 185 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 244
 |||||
 QY 61 AAAGAATGACTCTCTGAAGTGGAGGAG 87
 |||||
 Db 245 AAAGAATGACTCTCTGAAGTGGAGGAG 271
 |||||

RESULT 6
 BF791178 756 bp mRNA linear EST 12-JAN-2001
 LOCUS 602251278F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4338594 5',
 mRNA sequence.
 DEFINITION
 ACCESSION BF791178
 VERSION BF791178.1 GI:12096232
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 756)
 AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1214 row: 1 column: 19
 High quality sequence stop: 619.
 Location/Qualifiers
 1..756
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4338594"
 /clone_lib="NIH MGC 81"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site: 1: SfiI (ggccgctcgcc); Site 2: SfiI
 (ggccattatggcc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCGGCACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

FEATURES
 Source
 BASE COUNT 243 a 158 c 170 g 185 t
 ORIGIN
 Query Match 100.0%; Score 87; DB 12; Length 756;
 Best Local Similarity 100.0%; Pred. No. 8.1e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 60
 |||||
 Db 191 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 250
 |||||
 QY 61 AAAGAATGACTCTCTGAAGTGGAGGAG 87
 |||||
 Db 251 AAAGAATGACTCTCTGAAGTGGAGGAG 277
 |||||

RESULT 7
 BF671987 793 bp mRNA linear EST 21-DEC-2000
 LOCUS 601650456F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934333 5',
 mRNA sequence.
 DEFINITION
 ACCESSION BF126275
 VERSION BF126275
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 826)
 AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

DEFINITION
 ACCESSION BF671987
 VERSION BF671987.1 GI:11945882
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 793)
 AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLCM1141 row: 0 column: 02
 High quality sequence stop: 665.
 Location/Qualifiers
 1..793
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4293721"
 /clone_lib="NIH MGC 81"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: muscle (skeletal); Vector: pDNR-LIB
 (Clontech); Site: 1: SfiI (ggccgctcgcc); Site 2: SfiI
 (ggccattatggcc); 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAGAGCGGCGGCGGCACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

BASE COUNT 251 a 164 c 175 g 203 t
 ORIGIN
 Query Match 100.0%; Score 87; DB 12; Length 793;
 Best Local Similarity 100.0%; Pred. No. 8.2e-17;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 60
 |||||
 Db 192 GCAATATCAATATTCCTCAATGGAGGCTTTTCGGCCAGGAGCAGGTCAACCCCCAGAAGA 251
 |||||
 QY 61 AAAGAATGACTCTCTGAAGTGGAGGAG 87
 |||||
 Db 252 AAAGAATGACTCTCTGAAGTGGAGGAG 278
 |||||

RESULT 8
 BF126275 826 bp mRNA linear EST 24-OCT-2000
 LOCUS 601650456F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:3934333 5',
 mRNA sequence.
 DEFINITION
 ACCESSION BF126275
 VERSION BF126275
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 826)
 AUTHORS NIH-MGC http://imgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW774 row: p column: 14
High quality sequence stop: 651.
Location/Qualifiers

FEATURES
source

1. 826
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3934333"
/clone_lib="NIH MGC 76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgcctggcc); Site 2: SfiI (ggcgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."

BASE COUNT 280 a 156 c 188 g 202 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 826;
Best Local Similarity 100.0%; Pred. NO. 8.3e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATATGGAGCCTTCGGCCAGGAGCAGGTCAACCCCCAGAGA 60
|||||
DB 139 GCAATATCAATATTCATATGGAGCCTTCGGCCAGGAGCAGGTCAACCCCCAGAGA 198
|||||

QY 61 AAGAATGACTCTCTGAAGTGGAGGAG 87
|||||
DB 199 AAGAATGACTCTCTGAAGTGGAGGAG 225
|||||

RESULT 9
BF672902
LOCUS
DEFINITION 602152759F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293878 5', mRNA linear EST 21-DEC-2000
RNA sequence.

ACCESSION BF672902
VERSION BF672902.1 GI:11946797
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 855)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW142 row: e column: 15
High quality sequence stop: 687.
Location/Qualifiers

FEATURES
source

1. 879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4289735"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgcctggcc); Site 2: SfiI (ggcgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 270 a 159 c 210 g 216 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 855;
Best Local Similarity 100.0%; Pred. NO. 8.4e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATATGGAGCCTTCGGCCAGGAGCAGGTCAACCCCCAGAGA 60
|||||
DB 158 GCAATATCAATATTCATATGGAGCCTTCGGCCAGGAGCAGGTCAACCCCCAGAGA 217
|||||

QY 61 AAGAATGACTCTCTGAAGTGGAGGAG 87
|||||
DB 218 AAGAATGACTCTCTGAAGTGGAGGAG 244
|||||

RESULT 10
BF575112
LOCUS
DEFINITION 602134792F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4289735 5', mRNA linear EST 12-DEC-2000
RNA sequence.

ACCESSION BF575112
VERSION BF575112.1 GI:11648824
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCW131 row: h column: 24
High quality sequence stop: 595.
Location/Qualifiers

FEATURES
source

1. 879
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4289735"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgcctggcc); Site 2: SfiI (ggcgcctggcc); 5' and 3' adaptors were used in cloning as follows: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 284 a 147 c 239 g 208 t 1 others
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 879;
Best Local Similarity 100.0%; Pred. No. 8.5e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 190 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 249

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 250 AAAGAATGTACTCTGAGTGGAGGAG 276

RESULT 11
BF672126
LOCUS
DEFINITION 602152580F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293763 5',
ACCESSION BF672126
VERSION BF672126
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW141 row: p column: 20
High quality sequence stop: 621.
Location/Qualifiers
1. .906

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4293763"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfII (ggccgctcgcc); Site 2: SfII (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 285 a 208 c 200 g 213 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 906;
Best Local Similarity 100.0%; Pred. No. 8.5e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 190 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 249

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 250 AAAGAATGTACTCTGAGTGGAGGAG 276

RESULT 11
BF672126
LOCUS
DEFINITION 602152580F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293763 5',
ACCESSION BF672126
VERSION BF672126
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW141 row: p column: 20
High quality sequence stop: 621.
Location/Qualifiers
1. .906

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4293763"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfII (ggccgctcgcc); Site 2: SfII (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 159 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 218

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 219 AAAGAATGTACTCTGAGTGGAGGAG 245

RESULT 12
BF790243
LOCUS
DEFINITION 602249777F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
ACCESSION BF790243
VERSION BF790243
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW187 row: i column: 19
High quality sequence stop: 626.
Location/Qualifiers
1. .909

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328154"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfII (ggccgctcgcc); Site 2: SfII (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 279 a 200 c 212 g 218 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 8.6e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 228 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 287

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 288 AAAGAATGTACTCTGAGTGGAGGAG 314

RESULT 13
AU083330
LOCUS
DEFINITION 602249777F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
ACCESSION AU083330
VERSION AU083330
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW187 row: i column: 19
High quality sequence stop: 626.
Location/Qualifiers
1. .909

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328154"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfII (ggccgctcgcc); Site 2: SfII (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 279 a 200 c 212 g 218 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 8.6e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 228 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 287

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 288 AAAGAATGTACTCTGAGTGGAGGAG 314

RESULT 13
AU083330
LOCUS
DEFINITION 602249777F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
ACCESSION AU083330
VERSION AU083330
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW187 row: i column: 19
High quality sequence stop: 626.
Location/Qualifiers
1. .909

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328154"
/clone_lib="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site 1: SfII (ggccgctcgcc); Site 2: SfII (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCCATTTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 279 a 200 c 212 g 218 t
ORIGIN

Query Match 100.0%; Score 87; DB 12; Length 909;
Best Local Similarity 100.0%; Pred. No. 8.6e-17;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 60
Db 228 GCAATATCAATATTCATGAGGAGCCTTTCGCCAGGAGCAGGTCAACCCCGAGAGA 287

QY 61 AAAGAATGTACTCTGAGTGGAGGAG 87
Db 288 AAAGAATGTACTCTGAGTGGAGGAG 314

RESULT 13
AU083330
LOCUS
DEFINITION 602249777F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
ACCESSION AU083330
VERSION AU083330
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCW187 row: i column: 19
High quality sequence stop: 626.
Location/Qualifiers
1. .909

DEFINITION AU083330 Guinea pig organ of Corti pKF3 library Cavia porcellus
CDNA clone CLN11003, mRNA sequence.

ACCESSION AU083330
VERSION AU083330.1 GI:9988038
KEYWORDS EST.
SOURCE domestic guinea pig.
ORGANISM Cavia porcellus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.
1 (bases 1 to 261)
REFERENCE Nakaajima, T., Oshima, T., Wada, H., Ikeda, K. and Takasaka, T.
AUTHORS Characterization of novel and identified genes in guinea pig organ
TITLE of Corti cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Takaya Nakajima
Micromachine laboratory, Dept. of Mech. Eng.
Tohoku University
Aoba-yama 01, Sendai, Miyagi 980-8579, Japan
Tel: 81-22-217-4048
Fax: 81-22-217-6939
Email: nakajima@wadalab.mech.tohoku.ac.jp.

FEATURES
source
Location/Qualifiers
1..261
/organism="Cavia porcellus"
/strain="Hartley"
/db_xref="taxon:10141"
/clone="CLN11003"
/clone_lib="Guinea pig organ of Corti pKF3 library"
/tissue_type="organ of Corti"
/dev_stage="young adult"
BASE COUNT 81 a 68 c 60 g 52 t
ORIGIN
Query Match 90.8%; Score 79; DB 9; Length 261;
Best Local Similarity 94.3%; Pred. No. 1.8e-14;
Matches 82; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGAGGCTTCGGCCAGGAGGAGTCAACCCCGCAGAGA 60
Db 140 GCAATATCAATATTCATGGAGGCTTCGGCCAGGAGGAGTCAACCCCGCAGAGA 199

QY 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87
Db 200 AAAGAATGTACTCTCTGAAGTGGAGGAG 226

RESULT 14
BF790636
LOCUS
DEFINITION BF790636 867 bp mRNA linear EST 12-JAN-2001
602250354F1 NIH_MGC_81 Homo sapiens CDNA clone IMAGE:4337821 5',
mRNA sequence.

ACCESSION BF790636
VERSION BF790636.1 GI:12095690
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 867)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1212 row: 1 column: 14
High quality sequence stop: 514.

FEATURES
source
Location/Qualifiers
1..867
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4337821"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site_1: SfiI (ggcgcttcggcc); Site_2: SfiI
(ggcattatggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGCCATTATGCCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
BASE COUNT 248 a 165 c 244 g 210 t
ORIGIN
Query Match 87.4%; Score 76; DB 12; Length 867;
Best Local Similarity 98.9%; Pred. No. 2.3e-13;
Matches 87; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCAATATCAATATTCATGGAGGCTTCGGCCAGGAGGAGTCAACCCCGCAGAGA 60
Db 281 GCAATATCAATATTCATGGAGGCTTCGGCCAGGAGGAGTCAACCCCGCAGAGA 340

QY 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87
Db 341 AAAGAATGTACTCTCTGAAGTGGAGGAG 368

RESULT 15
AA389647
LOCUS
DEFINITION AA389647 394 bp mRNA linear EST 23-APR-1997
M104 Fetal heart, Lambda ZAP Express Homo sapiens cDNA 5', mRNA
sequence.

ACCESSION AA389647
VERSION AA389647.1 GI:2042633
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
REFERENCE Huang, D.M., Fung, Y.W., Wang, R.X., Laurensen, C.M., Ng, S.H., Lam
AUTHORS W.Y., Teui, K.W., Fung, K.P., Waye, M., Lee, C.Y. and Liew, C.C.
TITLE Analysis of expressed sequence tags from a fetal human heart cDNA
library
JOURNAL Genomics 30 (2), 293-298 (1995)
MEDLINE 96163883
COMMENT Contact: Kok LDS
Department of Biochemistry
The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong
Fax: 26035123
Email: ldskok@spider.net.hk
Seq primer: Forward Primer.
Location/Qualifiers
1..394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Fetal heart, Lambda ZAP Express"
/lab_host="E. coli"
/note="Vector: Lambda ZAP; Site 1: EcoRI; Site 2: XhoI;
The library is a gift from Prof. C.C. Liew, University of
Toronto, Ontario, Canada. mRNA was purified from human
fetal hearts (8-10 wks). cDNA was synthesized using a
XhoI-Oligo dT adaptor primer. EcoRI adaptors were ligated,
followed by digestion with XhoI, for directional cloning
into predigested lambda ZAP Express."

```

BASE COUNT      126 a      93 c      94 g      81 t
ORIGIN

Query Match      86.2%; Score 75; DB 9; Length 394;
Best Local Similarity 98.9%; Pred. No. 3.7e-13;
Matches 86; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GCAAATATCAATATTCCAATGGGAGCCTTTCCGCCAGGAGCAGGTCAACCCCGAGAAGA 60
   |||||||
Db 168 GCAAATATCAATATTCCAATGGGAGCCTTTCCGCCAGGAGCAGGTCAA-CCCCAGAAGA 226
   |||||||

QY 61 AAAGAATGTACTCTGTAAGTGGAGGAG 87
   |||||||
Db 227 AAAGAATGTACTCTGTAAGTGGAGGAG 253
   |||||||

```

Search completed: April 15, 2003, 22:02:32
 Job time : 181.443 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 18:42:01 ; Search time 5.14773 Seconds
(without alignments)
5183.040 Million cell updates/sec

Title: US-09-647-019-8

Perfect score: 87
Sequence: 1 gcaaatatcaatattcccaat.....gtactcctgaagtggaggag 87

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	909	4	US-09-484-970B-111
2	24.6	28.3	1877	4	US-09-780-173A-10
3	24.6	28.3	40328	3	US-08-742-185-102
4	24.6	28.3	43795	3	US-08-742-185-101
5	24.2	27.8	6924	1	US-08-015-973-2
6	24.2	27.8	6924	2	US-08-448-164-2
7	24.2	27.8	7941	4	US-09-816-703A-1
8	23.8	27.4	1070	4	US-09-470-443-7
9	23.8	27.4	1817	4	US-09-594-193-3
10	23.8	27.4	4112	4	US-09-221-017B-191
11	23.6	27.1	254	4	US-09-103-875-11
12	23.6	27.1	289	3	US-08-866-340-7
13	23.6	27.1	3404	4	US-09-453-702B-94
14	23.4	26.9	1752	4	US-09-022-949-1
15	23.4	26.9	1755	2	US-08-317-785-1
16	23.4	26.9	2254	4	US-08-552-369-1
17	23.4	26.9	5049	1	US-08-336-345-1
18	23.4	26.9	5049	1	US-08-336-345-2
19	23.4	26.9	5049	2	US-08-647-655-1
20	23.4	26.9	5049	2	US-08-647-655-2
21	23.4	26.9	9919	3	US-08-880-179-1
22	23.2	26.7	13187	4	US-09-422-936-61
23	23	26.4	1071	4	US-08-887-534A-71
24	23	26.4	2310	4	US-09-548-372D-54
25	23	26.4	2310	4	US-09-548-367D-54
26	23	26.4	2316	4	US-09-548-372D-58
27	23	26.4	2316	4	US-09-548-367D-58

c	28	23	26.4	6733	4	US-09-124-541-2	Sequence 2, Appli
c	29	22.8	26.2	1724	4	US-09-117-860-36	Sequence 36, Appl
c	30	22.6	26.0	1062	2	US-08-611-757-33	Sequence 33, Appl
c	31	22.6	26.0	1062	5	PCT-US95-05980-33	Sequence 33, Appl
	32	22.6	26.0	2375	1	US-08-369-796-9	Sequence 9, Appli
	33	22.6	26.0	2375	2	US-08-852-091-9	Sequence 9, Appli
	34	22.6	26.0	2375	2	US-08-820-754-9	Sequence 9, Appli
	35	22.6	26.0	2375	3	US-08-956-652-9	Sequence 9, Appli
	36	22.6	26.0	2375	3	US-08-956-652-9	Sequence 9, Appli
	37	22.6	26.0	2375	3	US-08-948-547-9	Sequence 9, Appli
	38	22.6	26.0	2375	5	PCT-US95-17025-9	Sequence 9, Appli
	39	22.4	25.7	824	4	US-08-998-416-339	Sequence 339, App
c	40	22.4	25.7	1817	1	US-08-459-586-8	Sequence 8, Appli
c	41	22.4	25.7	1817	2	US-08-282-696-8	Sequence 8, Appli
c	42	22.4	25.7	3762	4	US-09-318-448-17	Sequence 17, Appli
c	43	22.4	25.7	5924	1	US-08-459-586-6	Sequence 6, Appli
c	44	22.4	25.7	5924	2	US-08-282-696-6	Sequence 6, Appli
c	45	22.4	25.7	70000	4	US-09-851-896-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-484-970B-111
; Sequence 111, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 111
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 021656.2CB1
US-09-484-970B-111

Query Match	100.0%	Score 87;	DB 4;	Length 909;
Best Local Similarity	100.0%	Pred. No. 2.6e-22;		
Matches	87;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	GCAATATCAATATTC	CAATGGAGCCTTT	CGGCCAGGAGCAGGTCAACCCCGAGAAGA 60
DB	267	GCAATATCAATATTC	CAATGGAGCCTTT	CGGCCAGGAGCAGGTCAACCCCGAGAAGA 326
QY	61	AAAGAATGTA	CTCCTGAAGTGGAGGAG 87	
DB	327	AAAGAATGTA	CTCCTGAAGTGGAGGAG 353	
RESULT 2				
US-09-780-173A-10				
; Sequence 10, Application US/09780173A				
; Patent No. 6455307				
GENERAL INFORMATION:				
* APPLICANT: Robert McKay				
APPLICANT: Susan M. Freier				
APPLICANT: Jacqueline Wyatt				
TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA PRIME EXPRESSION				
FILE REFERENCE: RTS-0165				
CURRENT APPLICATION NUMBER: US/09/780,173A				
CURRENT FILING DATE: 2001-02-08				
NUMBER OF SEQ ID NOS: 95				
SEQ ID NO 10				

LENGTH: 1877
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (344)...(1396)
US-09-780-173A-10

Query Match 28.3%; Score 24.6; DB 4; Length 1877;
Best Local Similarity 76.9%; Pred. No. 10;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 11 ATATTCGAATGGAGCCCTTTCGCCAGGAGCGTCAAC 49
Db 1022 ATATTCGAATGGAGCCCTTTCGCCAGGAGCGTCAAC 1060

RESULT 3

US-08-742-185-102/c
Sequence 102, Application US/08742185
Patent No. 6020476
GENERAL INFORMATION:
APPLICANT: Page, David C.
APPLICANT: Reijo, Renee
APPLICANT: Saxena, Richa
APPLICANT: Hawkins, Trevor
APPLICANT: Reeve, Mary Pat
TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,185
FILING DATE: 30-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/690,734
FILING DATE: 31-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/310,429
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-07A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 40328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-742-185-102

Query Match 28.3%; Score 24.6; DB 3; Length 40328;
Best Local Similarity 59.2%; Pred. No. 26;
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 3 AAATATCAATATTCGAATGGAGCCCTTTCGCCAGGAGCGTCAACCCCGAGAGAAA 62
Db 6990 AACTCTGGAGTTTCAAGCGAGTCTTTGTCTACCCAGGTGAGAACTCAAGTTTAAA 6931

Db 34866 AACTCTGGAGTTTCAAGCGAGTCTTTGTCTACCCAGGTGAGAACTCAAGTTTAAA 34807
Qy 63 AGAATGTACTC 73
Db 34806 AAAATATTTTC 34796

RESULT 4

US-08-742-185-101/c
Sequence 101, Application US/08742185
Patent No. 6020476
GENERAL INFORMATION:
APPLICANT: Page, David C.
APPLICANT: Reijo, Renee
APPLICANT: Saxena, Richa
APPLICANT: Hawkins, Trevor
APPLICANT: Reeve, Mary Pat
TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,185
FILING DATE: 30-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/690,734
FILING DATE: 31-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/310,429
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-07A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 43795 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-742-185-101

Query Match 28.3%; Score 24.6; DB 3; Length 43795;
Best Local Similarity 59.2%; Pred. No. 27;
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 3 AAATATCAATATTCGAATGGAGCCCTTTCGCCAGGAGCGTCAACCCCGAGAGAAA 62
Db 6990 AACTCTGGAGTTTCAAGCGAGTCTTTGTCTACCCAGGTGAGAACTCAAGTTTAAA 6931

Qy 63 AGAATGTACTC 73

Db 6930 AAAATATTTTC 6920

RESULT 5

US-08-015-973-2
Sequence 2, Application US/08015973

; Patent No. 5604094
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,973
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6924 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6924
; US-08-015-973-2

Query Match 27.8%; Score 24.2; DB 1; Length 6924;
Best Local Similarity 62.3%; Pred. No. 21;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 12 TATTCCAATGGAGCCTTTCGGCCAGGAGGTCAACCCCGAGAGAAAGAAATGTAC 71
DB 3875 TGTTCACAAACGGCCAAATTTGGAGATTAAACAGGCCCATCCCGAGGAGGATGTAT 3934
QY 72 T 72
DB 3935 T 3935

RESULT 6
US-08-448-164-2
; Sequence 2, Application US/08448164
; Patent No. 5925536
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,164
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/015,973
; FILING DATE: 10-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6924 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6924
; US-08-448-164-2

Query Match 27.8%; Score 24.2; DB 2; Length 6924;
Best Local Similarity 62.3%; Pred. No. 21;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 12 TATTCCAATGGAGCCTTTCGGCCAGGAGGTCAACCCCGAGAGAAAGAAATGTAC 71
DB 3875 TGTTCACAAACGGCCAAATTTGGAGATTAAACAGGCCCATCCCGAGGAGGATGTAT 3934
QY 72 T 72
DB 3935 T 3935

RESULT 7
US-09-816-703A-1
; Sequence 1, Application US/09816703A
; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 AGY
; CURRENT APPLICATION NUMBER: US/09/816,703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 7941
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (148)..(7092)
; US-09-816-703A-1

Query Match 27.8%; Score 24.2; DB 4; Length 7941;
Best Local Similarity 62.3%; Pred. No. 22;
Matches 38; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 12 TATTCCAATGGAGCCTTTCGGCCAGGAGGTCAACCCCGAGAGAAAGAAATGTAC 71
DB 3875 TGTTCACAAACGGCCAAATTTGGAGATTAAACAGGCCCATCCCGAGGAGGATGTAT 3934
QY 72 T 72
DB 3935 T 3935

Query Match 27.1%; Score 23.6; DB 4; Length 3404;
Best Local Similarity 58.6%; Pred. No. 28;

```
Matches 41; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 16 CCAATGGAGCCTTCGGCCAGGAGGTCACACCCCAAGAAAGATGACTCCT 75
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2116 CCATGGCTCAATTCGGGCGGACGACACCTCCACCCCAAGATTAAATGAATGTGATCTA 2175
QY 76 GAAGTGGAGG 85
    |||||
Db 2176 ATATCTGAGG 2185

RESULT 14
US-09-022-949-1
; Sequence 1, Application US/09022949
; Patent No. 6187759
; GENERAL INFORMATION:
; APPLICANT: Tarpey, Ian
; APPLICANT: Greenwood, Neil
; TITLE OF INVENTION: Canine parvovirus DNA vaccination
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 6187759el Patent Dept.
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30(EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/022,949
; FILING DATE: 11-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-948-7400
; TELEFAX: 301-948-9751
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1752 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-09-022-949-1

Query Match 26.9%; Score 23.4; DB 4; Length 1752;
Best Local Similarity 63.2%; Pred. No. 27;
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 22 GGAGCCTTTCGGCCAGGAGGTCACACCCCAAGAAAGATGACTCCTGAA 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 GGAGCAGTTACACGAGCGTGTCTCAGAAATGAAGAGCTACAGGA 66

RESULT 15
US-08-317-785-1
; Sequence 1, Application US/08317785
; Patent No. 5882652
; GENERAL INFORMATION:
; APPLICANT: Elena Cortes Valds
; APPLICANT: Carmen Vela Olmo
; APPLICANT: Jos Ignacio Casal Alvarez
; TITLE OF INVENTION: "PROCESS FOR PRODUCING A
; TITLE OF INVENTION: SUBUNIT VACCINE AGAINST THE CANINE PARVOVIRUS AND OTHER
; TITLE OF INVENTION: RELATED VIRUSES"
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: HARRISON & EGBERT
; STREET: 1018 Preston Street, Suite 100
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 5.25 inch., 1.2 MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS (version 3.30)
; SOFTWARE: WordPerfect5.1 (WP5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,785
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/983,577
; FILING DATE: 26-JAN-1993
; APPLICATION NUMBER: PCT/ES92/00031
; FILING DATE: 26-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: John S. Egbert
; REGISTRATION NUMBER: 30627
; REFERENCE/DOCKET NUMBER: 11521
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 223-4034
; TELEFAX: (713) 223-4873
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1755 base pairs (585 amino acids)
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: Yes
; ANTI-SENSE: NO
; FRAGMENT TYPE: Complete VP2 sequence of CPV
; ORIGINAL SOURCE:
; ORGANISM: Canine Parvovirus
; STRAIN: CPV-N
; IMMEDIATE SOURCE:
; LIBRARY: Genomic
; PUBLICATION INFORMATION:
; AUTHORS:
; AUTHORS: Reed, A.P.
; AUTHORS: Jones, E.V.
; AUTHORS: Miller, T.J.
; TITLE: "NUCLEOTIDE SEQUENCE AND GENOME ORGANIZATION
; TITLE: OF CANINE PARVOVIRUS"
; JOURNAL: Journal of Virology
; VOLUME: 62
; ISSUE: 1
; PAGES: 266-276
; DATE: JAN, 1988
; RELEVANT RESIDUES IN SEQ ID NO: From 1 to 1755
US-08-317-785-1

Query Match 26.9%; Score 23.4; DB 2; Length 1755;
Best Local Similarity 63.2%; Pred. No. 27;
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 22 GGAGCCTTTCGGCCAGGAGGTCACACCCCAAGAAAGATGACTCCTGAA 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 GGAGCAGTTACACGAGCGTGTCTCAGAAATGAAGAGCTACAGGA 66

Search completed: April 15, 2003, 22:52:34
Job time : 26.1477 secs
```

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 19:13:41 ; Search time 8.69318 seconds
(without alignments)
8778.558 Million cell updates/sec

Title: US-09-647-019-8

Perfect score: 87

Sequence: 1 gcaaatatcaattccaat.....gtactcctgaaggaggag 87

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	824	10	US-09-880-192-4
2	74.2	85.3	466	10	US-09-960-352-5216
3	29.4	33.8	507	10	US-09-783-590-8009
4	28	32.2	1497	10	US-09-779-144A-6
5	26.4	30.3	211	10	US-09-974-300-1137
6	25.8	29.7	473	10	US-09-864-761-14995
7	25.6	29.4	661	9	US-09-989-920-147
8	25.6	29.4	1897	9	US-09-989-920-148
9	25.6	29.4	2350	10	US-09-880-192-36
10	25.2	29.0	41104	10	US-09-816-685-3
11	25	28.7	473	10	US-09-864-761-14346
12	24.6	28.3	276	10	US-09-923-876-4026
13	24.4	28.0	3829	12	US-10-002-600-3
14	24.2	27.8	278	10	US-09-878-574-8855
15	24.2	27.8	373	9	US-09-796-692-4296
16	24.2	27.8	407	10	US-09-517-800A-269
17	24.2	27.8	4883	9	US-10-118-513A-5
18	24.2	27.8	7215	9	US-10-118-513A-13
19	24.2	27.8	7215	9	US-10-118-513A-15

ALIGNMENTS

RESULT 1

US-09-880-192-4

; Sequence 4, Application US/09880192

; Patent No. US20020077470A1

; GENERAL INFORMATION:

; APPLICANT: Walker, Michael G.

; APPLICANT: Volkmutch, Wayne

; APPLICANT: Klingler, Tod M.

; APPLICANT: Azimzal, Yalda

; * TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION

; FILE REFERENCE: PB-0009-1 CIP

; CURRENT APPLICATION NUMBER: US/09/880,192

; CURRENT FILING DATE: 2001-06-12

; NUMBER OF SEQ ID NOS: 62

; SOFTWARE: PERL Program

; SEQ ID NO 4

; TYPE: DNA

; LENGTH: 824

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; OTHER INFORMATION: Incyte ID No. US20020077470A1 3601719CBI

US-09-880-192-4

Query Match

Best Local Similarity 100.0%; Score 87; DB 10; Length 824;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATTCGAGGAGCTTTTCGCCAGGAGCAGGTCAACCCCAAGA 60

Db 263 GCAATATCAATATTCATTCGAGGAGCTTTTCGCCAGGAGCAGGTCAACCCCAAGA 322

QY 61 AAGAATGTACTCTCTGAAGTGGAGGAG 87

Db 323 AAGAATGTACTCTCTGAAGTGGAGGAG 349

RESULT 2

US-09-960-352-5216

; Sequence 5216, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

```
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Batt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5216
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 23-BOVMS1-014-Q1-E1-F3
US-09-960-352-5216

Query Match      85.3%; Score 74.2; DB 10; Length 466;
Best Local Similarity 90.8%; Pred. No. 2e-18;
Matches 79; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      1  GCAATATCATATTCGAATGGAGCCTTCGGCCAGGAGGAGTCAACCCCGAGAGA 60
Db      148  GCAATATCATATTCGAATGGAGCCTTCGGCCAGGAGGAGTCAACCCCGAGAGA 207

QY      61  AAAGAATGCTACTCTCGAATGGAGGAG 87
Db      208  AAAGAATGCTACTCGGAGATAGAGGAG 234

RESULT 3
US-09-783-590-8009
; Sequence 8009, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: P0-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8009
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (30)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (115)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (124)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (150)
; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature
; LOCATION: (229)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (233)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (254)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (258)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (261)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (266)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (292)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (293)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (296)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (301)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (302)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (315)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (329)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (345)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (367)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (370)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (376)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (383)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (388)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (403)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (407)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

LOCATION: (409)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (419)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (423)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (424)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (425)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (434)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (443)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (445)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (457)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (458)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (482)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (486)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (489)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (490)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (501)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (502)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (504)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-8009

Query Match 33.8%; Score 29.4; DB 10; Length 507;
Best Local Similarity 60.0%; Pred. Nq. 0.16;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 5 ATATCAATATTCATCGGAGCCTTCGCGCAGGAGGAGGTCAACCCCGAGAGAAAAG 64
Db 144 ATTCTNTTTGAGATCGGAGCCTTAGTGCAAGTAAGAGACGACCTCAAAAAA 203
QY 65 AATGTACTCTGAAGTGGAG 84
Db 204 GTTGCCAAATGAAGGGTG 223

RESULT 4
US-09-779-144A-6/c
Sequence 6, Application US/09779144A
Patent No. US20020148006A1
GENERAL INFORMATION:
APPLICANT: Nes, David W.
TITLE OF INVENTION: Transgenic Plants with Modified Sterol Compositions
FILE REFERENCE: 11899.0198.DVUS01 MOBT:198

CURRENT APPLICATION NUMBER: US/09/779,144A
CURRENT FILING DATE: 2001-02-08
PRIOR APPLICATION NUMBER: US 09/106,926
PRIOR FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 1497
TYPE: DNA
ORGANISM: Zea mays
US-09-779-144A-6

Query Match 32.2%; Score 28; DB 10; Length 1497;
Best Local Similarity 63.2%; Pred. No. 0.72; Indels 0; Gaps 0;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 12 TATTCAATGGGAGCCTTCGCGCAGGAGGAGGTCAACCCCGAGAGAAAAGATGTAC 71
Db 1329 TATGCCACTGAGAGGTTTATGGCGAAAGAGATTCCAAACCCCTAGAAACGATAACC 1270
QY 72 TCCTGAAG 79
Db 1269 TCAAGCAG 1262

RESULT 5
US-09-974-300-1137
Sequence 1137, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1137
LENGTH: 211
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1137

Query Match 30.3%; Score 26.4; DB 10; Length 211;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 4 AATATCAATATTCATCGGAGCCTTCGCGCAGGAGGAGGTCAACCCCGAGAGAAA 63
Db 106 AGATGGAGATTCGGATTGTCGCCGAACGCGCGGAACAGTCAAAACCGTCAGAGAAG 185
QY 64 GAATGTACTCTGAAGTGGAGGAG 87
Db 166 GAAGGTGATTTTGCCGATGAAGG 189

RESULT 6
US-09-864-761-14995
Sequence 1495, Application US/09864761
Patent No. US2002048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 14995
LENGTH: 473
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009487.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
US-09-864-761-14995

Query Match 29.7%; Score 25.8; DB 10; Length 473;
Best Local Similarity 58.4%; Pred. No. 3.5;
Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GCAAATATCAATATTCGAATGGGAGCCTTTTCGGCCAGAGCAGGTCAACCCCAAGA 60
Db 18 GAACCTATTAAAGTAGCAATTCGAGAAATTTGCTAGACGTAAATTAAGACACAAGAAT 77
QY 61 AAAGAATGTACTCTCTGA 77
Db 78 AAAGAATTAATCTTCTGAGA 94

RESULT 7

US-09-989-920-147/c
Sequence 147, Application US/09889920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto

APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 147
LENGTH: 661
TYPE: DNA
ORGANISM: Homo sapien
US-09-989-920-147

Query Match 29.4%; Score 25.6; DB 9; Length 661;
Best Local Similarity 57.5%; Pred. No. 4.6;
Matches 46; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GCAAATATCAATATTCGAATGGGAGCCTTTTCGGCCAGAGCAGGTCAACCCCAAGA 60
Db 287 GCAAGTCTGAAAAAGCAAGGGGATCTTTGGGGCTAACTTCGGGATCCCTGCACCTTTATG 228
QY 61 AAAGAATGTACTCTCTGAAGT 80
Db 227 TAAGAATGTAACCTGGAGT 208

RESULT 8

US-09-989-920-148/c
Sequence 148, Application US/09889920
Patent No. US20020172957A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
FILE REFERENCE: DEX-0291
CURRENT APPLICATION NUMBER: US/09/989,920
CURRENT FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/252,500
NUMBER OF SEQ ID NOS: 284
SOFTWARE: PatentIn version 3.1
SEQ ID NO 148
LENGTH: 1897
TYPE: DNA
ORGANISM: Homo sapien
US-09-989-920-148

Query Match 29.4%; Score 25.6; DB 9; Length 1897;
Best Local Similarity 57.5%; Pred. No. 6.2;
Matches 46; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GCAAATATCAATATTCGAATGGGAGCCTTTTCGGCCAGAGCAGGTCAACCCCAAGA 60
Db 996 GCAAGTCTGAAAAAGCAAGGGGATCTTTGGGGCTAACTTCGGGATCCCTGCACCTTTATG 937
QY 61 AAAGAATGTACTCTCTGAAGT 80
Db 936 TAAGAATGTAACCTGGAGT 917

RESULT 9

US-09-880-192-36
Sequence 36, Application US/09880192
Patent No. US20020077470A1
GENERAL INFORMATION:

APPLICANT: Walker, Michael G.
APPLICANT: Volkmut, Wayne
APPLICANT: Klingler, Tod M.
APPLICANT: Azimzai, Valda
TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
FILE REFERENCE: PB-0009-1 CIP
CURRENT APPLICATION NUMBER: US/09/880,192
CURRENT FILING DATE: 2001-06-12
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PERL Program
SEQ ID NO 36
LENGTH: 2350
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020077470A1 3434460CT1
US-09-880-192-36

Query Match 29.4%; Score 25.6; DB 10; Length 2350;
Best Local Similarity 57.5%; Pred. No. 6.6;
Matches 46; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCGAATGGAGCCTTTCGGCCAGGACAGGTCAACCCCGAGAAGA 60
|||||
Db 878 GCAAGTCTGAAGGCAAGGGGATCTTGGGGCTAACTTCGGGATCCCTGCACCTTTATG 937

QY 61 AAAGAATGTAATCTCTGAAGT 80

Db 938 TAAGATGTAAACCTGGAGT 957

RESULT 10
US-09-816-685-3/c
Sequence 3: Application US/09816685
Patent No. US20020053091A1
GENERAL INFORMATION:
APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
TITLE OF INVENTION: PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000856
CURRENT APPLICATION NUMBER: US/09/816,685
CURRENT FILING DATE: 2001-03-26
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 41104
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)_(41104)
OTHER INFORMATION: n = A,T,C or G
US-09-816-685-3

Query Match 29.0%; Score 25.2; DB 10; Length 41104;
Best Local Similarity 60.0%; Pred. No. 21;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCGAATGGAGCCTTTCGGCCAGGACAGGTCAACCCCGAGAAGA 60
Db 20755 GGACATTCGAATTTGACCTTGTTGCTTCTTGGCCAGATACTCTTCTCCCTCCCTAANA 20696

QY 61 AAAGAATGTA 70

Db 20695 GAATAAAGTA 20686

RESULT 11
US-09-761-14346/c
Sequence 14346, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/006666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14346
LENGTH: 473
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006337.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.71
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.73
US-09-864-761-14346

Query Match 28.7%; Score 25; DB 10; Length 473;
Best Local Similarity 58.9%; Pred. No. 7;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 5 ATATCAATATTCGAATGGAGCCTTTCGGCCAGGACAGGTCAACCCCGAGAAGAAG 64

Db 187 ATCTCAGGCTGCTTTTGAGAGGATCCCGCTGAGCCAGTTCCTGCCCTCCCTGATAAACA 128

QY 65 AATGTACTCCTGA 77

Db 127 AAGGCAAACTGA 115

RESULT 12
US-09-923-876-4026
; Sequence 4026, Application US/09923876
; Patent No. US2002013958A1
; GENERAL INFORMATION:
; APPLICANT: Kamigaki, Raghunath V.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 4026
; LENGTH: 276
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700454286H1
; NAME/KEY: unsure
; LOCATION: 174
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-4026

Query Match 28.3%; Score 24.6; DB 10; Length 276;
Best Local Similarity 70.2%; Pred. No. 8.4;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 10 AATATTCATATTCGAATGAGCGCTTTGGCCAGGAGGAGGTCACCCCGCCAG 56
Db 19 AATATTCATATTCGAATGAGCGCTTTGGCCAGGAGGAGGTCACCCCGCCAG 65

RESULT 13
US-10-002-600-3/c
; Sequence 3, Application US/10002600
; Patent No. US20020137077A1
; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocke, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 3829
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 238104.1
US-10-002-600-3

Query Match 28.0%; Score 24.4; DB 12; Length 3829;
Best Local Similarity 56.1%; Pred. No. 21;
Matches 46; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 2 CAATATCAATATTCGAATGAGCGCTTTGGCCAGGAGGAGGTCACCCCGCCAGAGAA 61
Db 2659 CAAAACACTAGTTAAAAATGTAACATTTTCAGCCATCTCTTGTAGATCACACAGACAA 2600

QY 62 AAGAATGTACTCTCTGAAGTGA 83
Db 2599 AAAAGGTAAACATTACATGAA 2578
RESULT 14
US-09-878-574-8855
; Sequence 8855, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J. D.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 8855
; LENGTH: 278
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701101759H1
US-09-878-574-8855

Query Match 27.8%; Score 24.2; DB 10; Length 278;
Best Local Similarity 71.1%; Pred. No. 12;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 24 AGCTTTTCGCGAGGAGGTCACCCCGCCAGAGAGAAAGATG 68
Db 6 AGTTTTTGAGCTGGCTCTGAACAAATCCCGAGAGAAAGATG 50

RESULT 15
US-09-796-692-4296/c
; Sequence 4296, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04


```
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4296
; LENGTH: 373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (147)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (232)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (267)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (351)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-4296
```

```
Query Match      27.8%; Score 24.2; DB 9; Length 373;
Best Local Similarity 58.6%; Pred. No. 13;
Matches 41; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY  18  AATGGAGCCTTTGGCCAGGACGAGTCAACCCCCAGAGAAAGATGTACTCCTGA 77
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   208  AACAGACGCCATCCTACACCTATCAGCCCTGCTGCCAGGAGAACAGACACTCCTGG 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY  78  AGTGGAGGAG 87
      ||| ||| |||
Db   148  TTGGATGGG 139
```

Search completed: April 16, 2003, 01:14:27
Job time : 18.6932 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:26:41 ; Search time 23.2159 Seconds
(without alignments)
8439.207 Million cell updates/sec

Title: US-09-647-019-8
Perfect score: 87
Sequence: 1 gcaaatatcaatttccaat.....gtactctgaagtggaggag 87

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 115999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	N Geneseq 101002:*
1:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	428	21	AAC01483 Human secreted pro
2	87	100.0	886	24	AAD27216 Human 66214 EST cl
3	87	100.0	887	20	AAK90904 cDNA encoding huma
4	66.2	76.1	778	20	AAK90903 cDNA encoding muri
5	28	32.2	1383	21	AAZ92614 Maize C-24 sterol
6	28	32.2	1497	20	AAV70839 Corn SMT cDNA. Ze
7	27	31.0	612	22	AAH71474 Human cervical can
8	27	31.0	42519	22	AAK81318 Human immune/haema
9	26.4	30.3	211	24	ABK73846 Bacillus lichenifo

C 10	26.4	30.3	891	22	AAH06869 Human cDNA clone (
C 11	26.2	30.1	11516	20	AAK20536 Polynucleotide seq
C 12	26.2	30.1	27435	20	AAK65476 Human immune/haema
C 13	26	29.9	1133	20	AAZ25118 Corn anthranilate
C 14	25.8	29.7	473	22	ABA43515 Human breast cell
C 15	25.8	29.7	473	22	ABA53965 Human foetal liver
C 16	25.8	29.7	473	22	AAI12262 Probe #2195 for ge
C 17	25.8	29.7	473	22	AAI102176 Probe #2167 used t
C 18	25.8	29.7	473	24	ABS02143 Human genome-deriv
C 19	25.8	29.7	830	20	AAZ16633 Human gene express
C 20	25.8	29.7	1641	21	AAK59785 Human secreted pro
C 21	25.6	29.4	409	22	AAK07822 Cervical cancer pr
C 22	25.6	29.4	1431	24	ABN07872 Human ovary specif
C 23	25.6	29.4	1667	20	AAZ42152 Human normal bladd
C 24	25.4	29.2	444	22	AAK08552 Human reproductive
C 25	25.4	29.2	2676	23	AAK78484 DNA encoding novel
C 26	25.2	29.0	10113	22	AAK32838 Human genomic DNA
C 27	25.2	29.0	11655	22	AAK32839 Human G-protein co
C 28	25.2	29.0	41104	24	AAK32660 Human foetal liver
C 29	25	28.7	473	22	ABA57946 Human bone marrow
C 30	25	28.7	473	22	AAK31665 Probe #6228 used t
C 31	25	28.7	473	22	AAI37542 Human genome-deriv
C 32	25	28.7	473	24	ABS06420 Human immune/haema
C 33	24.8	28.5	8095	22	AAK81626 Human prostate exp
C 34	24.8	28.5	8630	19	AAV23082 Arabidopsis thalia
C 35	24.6	28.3	446	23	ABV48465 Human prostate exp
C 36	24.6	28.3	1538	21	AAK45060 Arabidopsis thalia
C 37	24.6	28.3	1541	21	AAK33555 Arabidopsis thalia
C 38	24.6	28.3	1809	18	AAK84169 Arabidopsis thalia
C 39	24.6	28.3	13655	23	ABL10268 Arabidopsis thalia
C 40	24.6	28.3	40328	21	AAZ92584 Arabidopsis thalia
C 41	24.6	28.3	43795	21	AAZ92583 Arabidopsis thalia
C 42	24.4	28.0	2993	23	ABL27008 Arabidopsis thalia
C 43	24.4	28.0	3053	24	AAK99629 Arabidopsis thalia
C 44	24.2	27.8	322	22	AAH03784 Human cDNA clone (
C 45	24.2	27.8	373	21	AAA69554 Pinus radiata cycl

ALIGNMENTS

RESULT 1
AAC01483
ID AAC01483 standard; cDNA; 428 BP.
XX
AC AAC01483;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 1481.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
(GEST) GENSET.
XX
Dumas Milne Edwards J, Duclert A, Giordano J;
XX
WPI; 2000-500381/45.
XX
P-PSDB; AAG01477.
XX
New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX
obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 1481; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

CC vectors.

XX Sequence 428 BP; 123 A; 111 C; 104 G; 90 T; 0 other;

SQ

Query Match 100.0%; Score 87; DB 21; Length 428;

Best Local Similarity 100.0%; Pred. No. 4.4e-22;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGGAGCCTTCGGCCAGGAGCAGGTCAACCCCGAGA 60

Db 293 GCAATATCAATATTCATGGGAGCCTTCGGCCAGGAGCAGGTCAACCCCGAGA 352

QY 61 AAGAATGTACTCTCTGAAGTGGAGGAG 87

Db 353 AAGAATGTACTCTCTGAAGTGGAGGAG 379

RESULT 2

AAAD27216

ID AAD27216 standard; DNA; 886 BP.

AC AAD27216;

XX

XX 09-APR-2002 (first entry)

XX Human 66214 EST clone DNA.

XX Human; congestive heart failure; dilative cardiomyopathy; sudden death;

XX hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;

XX heart muscle disease; conduction disorder; coronary heart disease;

XX systemic arterial hypertension; pulmonary hypertension; endocarditis;

XX pulmonary heart disease; valvular heart disease; pericardial disease;

XX congenital heart disease; gene therapy; syncope; transgenic animal;

XX expressed sequence tag; EST; clone 66214; ds.

XX Homo sapiens.

XX

XX Key Location/Qualifiers

XX CDS 184..450

FT /*tag= a

FT /product= "Human 66214 protein"

FT misc_feature 298..588

FT /*tag= b

FT /note= "66214 cDNA fragment"

FT polyA_signal 857..862

FT /*tag= c

XX

XX W0200192567-A2.

XX

XX 06-DEC-2001.

XX

XX 30-MAY-2001; 2001WO-EP06165.

XX

XX 30-MAY-2000; 2000US-207400P.

XX

XX (MED1-) MEDIGENE AG.

XX Bunk D, Reuner B, Beck J, Henkel T;

XX WPI; 2002-122073/16.

XX P-PSDB; AAE16632.

XX Identifying a subject at risk for a heart disease e.g. congestive heart

XX failure, dilative cardiomyopathy, heart muscle disease, by quantifying

XX the polypeptide expressed by genes abnormally expressed in heart tissue

XX

XX Claim 2a; Fig 9b; 154pp; English.

XX The patent discloses novel target genes abnormally expressed in heart

XX tissues and their corresponding proteins. The invention also relates to

XX methods for assessing the expression level of these genes. The method

XX is used for testing the predisposition of mammals and preferably humans

XX for a heart disease or for an acute state of such a disease. It is also

XX useful to treat diseases of the heart such as congestive heart failure,

XX dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-

XX myopathy, specific heart muscle disease, rhythm and conduction disorders,

XX syncope and sudden death, coronary heart disease, systemic arterial

XX hypertension, pulmonary hypertension, pulmonary heart disease, valvular

XX heart disease, congenital heart disease, pericardial disease and

XX endocarditis. Sequences of the invention are also used in gene therapy.

XX A transgenic non-human mammal comprising the sequences of the invention

XX are useful for the development for medicaments for the treatments of

XX heart diseases. The present DNA sequence is expressed sequence tag

XX (EST) 66214 clone.

XX Sequence 886 BP; 278 A; 172 C; 191 G; 245 T; 0 other;

SQ

Query Match 100.0%; Score 87; DB 24; Length 886;

Best Local Similarity 100.0%; Pred. No. 5.7e-22;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGGAGCCTTCGGCCAGGAGCAGGTCAACCCCGAGA 60

Db 229 GCAATATCAATATTCATGGGAGCCTTCGGCCAGGAGCAGGTCAACCCCGAGA 288

QY 61 AAGAATGTACTCTCTGAAGTGGAGGAG 87

Db 289 AAGAATGTACTCTCTGAAGTGGAGGAG 315

RESULT 3

AAAX90904

ID AAX90904 standard; cDNA; 887 BP.

XX

XX AAX90904;

XX

XX 17-JAN-2000 (first entry)

XX

XX cDNA encoding human chisel (Csl) gene.

XX Chisel gene; Csl; EF-Hand protein super family; muscle development;

XX heart/skeletal muscle cell development; signalling pathway; regulation;

XX Xq21.3-q22; adaptive process; muscle homeostasis; skeletal myopathy;

XX detection; diagnosis; prophylaxis; treatment; cardiac hypertrophy;

XX muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;

XX Becker's myotonic dystrophy; heart failure; differentiation; exon;

XX gene therapy; transgenic animal; drug screening; ss.

XX Homo sapiens.

XX

XX Key Location/Qualifiers

XX exon 1..172

FT /*tag= a

FT /label= Exon_1

FT /note= "Corresponds to residues 19497-19327 of

FT human cosmid clone U228D4"

FT exon 173..229

FT /*tag= b

FT FT /label= Exon 2
 FT /note= "Corresponds to residues 15687-15631 of
 FT human cosmid clone U228D4"
 FT CDS 185...451
 FT /tag= c
 FT /product= "Chisel (Csl) protein"
 FT /note= "Expressed predominantly in heart muscles"
 FT FT 230...316
 FT FT /tag= d
 FT FT /label= Exon 3
 FT FT /note= "Corresponds to residues 5220-5134 of
 FT human cosmid clone U228D4"
 FT FT 317...465
 FT FT /tag= e
 FT FT /label= Exon 4
 FT FT /note= "Corresponds to residues 35384-35236 of
 FT human cosmid clone U112E8"
 FT FT 466...887
 FT FT /tag= f
 FT FT /label= Exon 5
 FT FT /note= "Corresponds to residues 4101-3680 of
 FT human cosmid clone U112E8"
 FT FT
 XX WO9950410-A1.
 XX
 XX
 XX
 PD 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WO-AU00220.
 XX 27-MAR-1998; 98AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GEO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX
 XX WPI: 1999-610852/52.
 DR P-PSDB; AAY28651.
 XX
 XX Isolated nucleic acids encoding chisel, used to develop products for
 FT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 FT muscular myopathies -
 XX
 XX Claim 8; Page 149-150; 157pp; English.
 XX
 XX The present sequence is the cDNA encoding the human chisel gene (Csl)
 CC that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.
 CC It is a member of the EF-Hand protein super family and is involved in
 CC signalling pathways. It is predominantly expressed in heart and skeletal
 CC muscles and is activated after the differentiation of cells. Csl
 CC functions in regulation aspects of differentiation or adaptive processes
 CC that maintain muscle homeostasis. This sequence can be used in the
 CC detection, diagnosis, prophylactic and therapeutic treatment of diseases
 CC such as those involving aberrant muscle cell development and functional
 CC activity. It is also used in the treatment of muscular and myotonic
 CC dystrophies, skeletal myopathies such as Duchenne muscular dystrophy and
 CC Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,
 CC myocardiitis, myofiber atrophy, etc. The Csl gene sequence can also be
 CC used in gene therapy, for the production of transgenic animals and for
 XX drug screening.
 XX
 XX Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;
 SQ
 Query Match 100.0%; Score 87; DB 20; Length 887;
 Best Local Similarity 100.0%; Pred. No. 5.7e-22;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GCNAATATCAATATTCGAATGGAGCCCTTCGCCAGGAGGAGTCAACCCCCCAGAGA 60
 Db 230 GCNAATATCAATATTCGAATGGAGCCCTTCGCCAGGAGGAGTCAACCCCCCAGAGA 289
 Oy 61 AAAGAATGCTACTCTCTGAAGTGGAGGAG 87

Db 290 AAAGAATGCTACTCTCTGAAGTGGAGGAG 316
 |||||
 RESULT 4
 AAX90903
 ID AAX90903 standard; cDNA; 778 BP.
 XX
 AC AAX90903;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE cDNA encoding murine chisel (Csl) gene.
 XX
 KW Chisel gene; Csl; EF-Hand protein super family; muscle development;
 KW heart/skeletal muscle cell development; signalling pathway; murine;
 KW X chromosome; regulation; adaptive process; muscle homeostasis;
 KW detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KW muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KW Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KW differentiation; gene therapy; transgenic animal; drug screening; ss.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH 199..456
 FT /tag= a
 FT /product= "Chisel (Csl) protein"
 FT /note= "Expressed especially in heart muscles"
 XX
 PN WO9950410-A1.
 XX
 XX 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WO-AU00220.
 XX 27-MAR-1998; 98AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 PA (GEO) GEN HOSPITAL CORP.
 PA (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Musaro A, Palmer SJ, Rosenthal NA;
 XX
 XX WPI: 1999-610852/52.
 DR P-PSDB; AAY28650.
 XX
 XX Isolated nucleic acids encoding chisel, used to develop products for
 FT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 FT muscular myopathies -
 XX
 XX Claim 3; Page 148; 157pp; English.
 XX
 XX The present sequence is the cDNA encoding the murine chisel gene (Csl)
 CC that is mapped to the mouse X chromosome. It is a member of the EF-Hand
 CC protein super family and is involved in signalling pathways. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocardiitis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 XX
 XX Sequence 778 BP; 231 A; 166 C; 179 G; 202 T; 0 other;
 SQ
 Query Match 76.1%; Score 66.2; DB 20; Length 778;
 Best Local Similarity 85.1%; Pred. No. 2.4e-14;
 Matches 74; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy	1	GC	AAATATCAATATTC	CAATGGGAGCCTTTTGGCCAGGAGCAGTCAACCCCCAGAAGA	60
Db	238	CG	AAATATCAATATTC	CAATGGGAGCCTTTTGGCCAGGAGCCTCCAGAGG	324
Qy	61	AA	GAAATCTACTCTCT	CTGAAGTCGAGGAG	87
Db	298	AA	GAGAGTACTCTCT	GAACTGAGGAG	324
RESULT 5					
AAZ92614/c					
ID	AAZ92614	standard;	CDNA;	1383 BP.	
XX					
AC	AAZ92614;				
DT	05-JUN-2000	(first entry)			
DE					
XX					
XX					
KW	C-24 sterol methyltransferase; ESMT-1;	maize;	endosperm;	alkylation;	
KW	sterol metabolism; ERG6 homologue; EST;	expressed sequence tag;			
KW	transgenic plant; pest resistance;	insect;	nematode;	fungus; ss.	
XX					
OS	Zea mays.				
XX					
Key		Location/Qualifiers			
PH	78..1112				
FT	/tag= a				
FT	/product= "Maize C-24 sterol methyltransferase ESMT-1"				
FT	WO200008190-A2.				
XX					
PN	17-FEB-2000.				
XX					
PD					
XX					
PF	02-AUG-1999;	99WO-US17456.			
XX					
PR	03-AUG-1998;	98US-0128339.			
XX					
PA	(ARIZ-) ARIZONA BOARD OF REGENTS.				
XX					
PI	Grebenok RJ;				
XX					
DR	WPI; 2000-205728/18.				
DR	P-PSDB; AAY80992.				
XX					
PT	DNA fragment for altering sterol metabolism in plants to give				
PT	protection against infestation by pests, encodes sterol				
PT	methyltransferase				
XX					
PS	Claim 1; Page 22-25; 25pp; English.				
XX					
CC	This sequence represents cDNA encoding maize endosperm C-24 sterol				
CC	methyltransferase ESMT-1. The cDNA encoding ESMT-1 was selected from a				
CC	EST (expressed sequence tag) database on the basis of its sequence				
CC	identity with the N-terminal portion of the yeast ERG6 protein. The				
CC	predicted amino acid sequence of ESMT-1 shows 66% similarity and 46%				
CC	identity with yeast ERG6, and has 75% and 37% identity to soybean and				
CC	Arabidopsis sterol methyltransferases, respectively. ESMT-1 catalyzes				
CC	alkylation of sterol intermediates at carbon 24. Nucleic acids encoding				
CC	ESMT-1 may be used in the generation of transgenic plants having altered				
CC	sterol metabolism. Sterols are essential for the synthesis and				
CC	maintenance of membranes of most eukaryotic cells; however, insects,				
CC	nematodes and some species of fungi do not produce their own sterols,				
CC	relying on plants for the completion of their life cycles. DNA encoding				
CC	ESMT-1 is useful in altering sterol metabolism in plants and protects				
CC	plants against infestation by pests such as insects, nematodes and some				
CC	fungi. Plants produced using ESMT-1 DNA are resistant to pest				
CC	infestation, without the need for chemical pesticides which may be toxic				
CC	to a wide range of species.				
XX					
XX	Sequence 1383 BP; 377 A; 276 C; 370 G; 360 T; 0 other.				

```

Query Match          32.2%; Score 28; DB 21; Length 1383;
Best Local Similarity 63.2%; Pred. No. 3.3;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps

QY 12 TATTCCAATGGGAGCCTTTTCGGCCAGGACAGGTCAACCCCCCAGAGAAAAGATGTAC 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1341 TATCCCACTGAGAGGTTTATGGCGAAAGAGTTCACACCCCCACCTAGAAAAGTAACC 1282
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 72 TCCTGAAG 79
    |||||
Db 1281 TCAAGCAG 1274

RESULT 6
AAV70839/c
ID AAV70839 standard; DNA; 1497 BP.
XX
AC AAV70839;
XX
DT 01-MAR-1999 (first entry)
XX
DE Corn SMT cDNA.
XX
KW S-adenosyl-L-methionine-delta24(25)-sterol methyl transferase;
KW SMT; phytosterol; transgenic plant; disease resistance;
KW insect resistance; pest resistance; stress tolerance;
KW crop protection; corn; maize; ds.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT CDS 66..1100
FT FT /*tag= a
FT polyA_site 1470..1497
FT FT /*tag= b
XX
FN WO9845457-A1.
XX
PD 15-OCT-1998.
XX
PF 24-DEC-1997; 97WO-US23495.
XX
PR 26-DEC-1996; 96US-0033923.
XX
PA (MONS ) MONSANTO CO.
XX
PI Nes WD;
XX
DR WPI; 1999-034606/03.
DR P-PSDB; AAW70838.
XX
PT New transgenic plants with altered sterol levels - obtained using
PT DNA comprising a coding sequence which encodes an enzyme which binds
PT a first sterol and produces a second sterol
XX
PS Example 6; Page 71-73; 97pp; English.
XX
CC This DNA molecule codes for the S-adenosyl-L-methionine-delta24(25)
CC sterol methyltransferase (SMT) enzyme (see AAW70838) of corn. SMT
CC catalyses the transfer of a methyl group from S-adenosyl-L-methionine
CC to the C-24 centre of the sterol side chain. Full-length corn SMT
CC cDNA was isolated from a commercial corn cDNA library by PCR
CC amplification. The invention relates to transgenic plants with
CC modified sterol biosynthetic pathways. Phytosterol compositions are
CC modulated in the transgenic plants to confer resistance to insects,
CC nematodes, pythaceous fungi and/or other environmental stresses
CC such as drought or cold, and/or to improve nutritional value, e.g.
CC by increasing levels of cholesterol-lowering sterols. SMT DNA
CC molecules (including antisense molecules) can be transformed into
CC plant cells, and plants (especially tomato, corn or soybean) having
CC altered sterol compositions are then regenerated.
XX
SQ Sequence 1497 BP; 413 A; 292 C; 394 G; 397 T; 1 other:

```

```

Query Match      32.2%; Score 28; DB 20; Length 1497;
Best Local Similarity 63.2%; Pred. No. 3.4;
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

OY 12 TATTCATGCGAGCTTCGCCAGGAGCAGGTCAACCCGCCAGAAAGAAATGTAC 71
Db 1329 TATGCCACTGAGAGGTTTATGCGGAAGAGGATTCCAGCCGCCACCTAGAAAGATAACC 1270

OY 72 TCCTGAAG 79
Db 1269 TCAGCAG 1262

RESULT 7
ID AAH71474 standard; cDNA; 612 BP.
XX AC AAH71474;
XX DT 19-SEP-2001 (first entry)
XX DE Human cervical cancer marker nucleic acid 2748.
XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO200142467-A2.
XX PD 14-JUN-2001.
XX PF 08-DEC-2000; 2000WO-US33312.
XX PR 08-DEC-1999; 99US-0169681.
XX PR 21-DEC-1999; 99US-0171350.
XX PR 14-MAR-2000; 2000US-0189315.
XX PR 12-MAY-2000; 2000US-0203791.
XX PR 09-JUN-2000; 2000US-0210600.
XX PR 21-JUL-2000; 2000US-0220114.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Deeds J, Berger A, Zhao X;
XX DR WPI; 2001-375006/39.
XX PT New isolated nucleic acid for diagnosing and treating cervical cancer
XX PT and for assessing and detecting compounds for treating the cancer -
XX PS Claim 1; Page 564-565; 1051pp; English.
XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded
XX CC polypeptides are useful: to assess if a patient is afflicted with
XX CC cervical cancer or has a pre-malignant condition; to monitor the
XX CC progression of cervical cancer or a premalignant condition in a patient;
XX CC and to select and/or assess the efficacy of a compound or therapy for
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX CC useful for gene therapy.
XX SQ Sequence 612 BP; 225 A; 61 C; 93 G; 232 T; 1 other;

Query Match      31.0%; Score 27; DB 22; Length 612;
Best Local Similarity 60.0%; Pred. No. 5.9;
Matches 45; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

OY 4 AATATCATATTCATTCGAGCGCTTCGCCAGGAGCAGGTCAACCCGCCAGAAAGAAA 63
Db 444 AAAAAAAGCCCAATTTGGCGCTTGGGCTAAAGAGGGGCCACCCCTCTCTTAAAAAA 503

OY 64 GAATGTACTCTCTGAA 78

```

```

Db 504 AARATTTCTCTTAA 518

RESULT 8
ID AAK81318
XX AC AAK81318 standard; DNA; 42519 BP.
XX AC AAK81318;
XX DT 07-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:36130.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 14-JUL-2000; 2000US-0217496.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 23-AUG-2000; 2000US-0227182.
XX PR 30-AUG-2000; 2000US-0227009.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 05-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 06-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 08-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.

```

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 36130; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 42519 BP; 10741 A; 10117 C; 9745 G; 11916 T; 0 other;
Query Match 31.0%; Score 27; DB 22; Length 42519;
Best Local Similarity 57.8%; Pred. No. 26;
Matches 48; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
QY 3 AATATTCATATTCCTGAGGAGGCTTTCGGCCAGGAGGAGGTCACACCCCAAGAAA 62
Db 6464 AATATCAAAAATTCGCTGGGCGATGGTGGCAGGAGGCTCAGGAGGCTGACACAGAGAA 6523
QY 63 AGAATGTACTCTGAAGTGGAGG 85
Db 6524 CGCTTGAAACCCAGAGGTGGAGG 6546
RESULT 9
ID ABK73846
XX ABK73846 standard; DNA; 211 BP.
AC ABK73846;
XX
DT 13-AUG-2002 (first entry)
XX Bacillus licheniformis genomic sequence tag (GST) #1137.
DE
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.

XX OS Bacillus licheniformis.
 XX PN WO200229113-A2.
 XX PD 11-APR-2002.
 XX PF 05-OCT-2001; 2001WO-US31437.
 XX PR 06-OCT-2000; 2000US-0680598.
 XX PR 27-MAR-2001; 2001US-279526P.
 XX PA (NOVO) NOVOZYMES BIOTECH INC.
 XX PA (NOVO) NOVOZYMES AS.
 XX PI Berka R, Clausen IG;
 XX DR WPI; 2002-416684/44.
 XX PT Monitoring differential expression of several genes in first Bacillus
 PT cell relative to expression of same genes in one or more second
 PT Bacillus cells, by using substrate containing Bacillus genomic
 PT sequenced tag array -
 XX PS Claim 4; SEQ ID NO 1137; 200pp; English.
 XX CC The invention describes a method of monitoring differential expression of
 CC genes in a first Bacillus cell relative to expression of the genes in
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
 CC isolated from Bacillus cells to a substrate containing array of Bacillus
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first Bacillus cell relative to expression of the same genes
 CC in one or more second Bacillus cells. The method is useful for monitoring
 CC global expression of several genes from a Bacillus cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which Bacillus cells adapt to changes in culture conditions, of the way
 CC environmental stress or other physiological provocation. Extensive
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 211 BP; 71 A; 33 C; 62 G; 45 T; 0 other;
 Query Match 30.3%; Score 26.4; DB 24; Length 211;
 Best Local Similarity 57.1%; Pred. No. 6.7;
 Matches 48; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
 QY 4 AATATCAATATTCGAATGGAGGCTTTCCGCCAGGAGCAGGTCAACCCCGCAGAGAAA 63
 DB 106 AAGATGAGATTCGATTTGCGCGAAACGCCGACACTCAAAACCGTCAGAAAGAC 165
 QY 64 GAATGTACTCCTGAATGGAGGAG 87
 DB 166 GAAGGTGATTTTCCCGATGAAGG 189
 RESULT 10
 AAH06869/c
 ID AAH06869 standard; cDNA; 891 BP.
 AC AAH06869;
 XX DT 26-JUN-2001 (first entry)
 XX

DE Human cDNA clone (5'-primer) SEQ ID NO:3704.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; es.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX full-length cDNAs -
 XX PS Claim 1; SEQ ID 3704; 2537pp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 891 BP; 195 A; 206 C; 232 G; 254 T; 4 other;
 Query Match 30.3%; Score 26.4; DB 22; Length 891;
 Best Local Similarity 66.7%; Pred. No. 11;
 Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 26 CCTTTCCGCCAGGAGCAGGTCAACCCCGCAGAGAAATGACTCTCTGAAG 79
 DB 879 CCATAGGCCACCAAGNAGNCAACCCCGCAGTAGCAAGAGCCCAATGAAG 826
 RESULT 11
 AAX20536/c
 ID AAX20536 standard; DNA; 11516 BP.
 XX AAX20536;
 AC AAX20536;
 XX DT 05-MAY-1999 (first entry)
 XX

XX Polynucleotide sequence from the genome of Treponema pallidum.
DE Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX enzyme production; ds.
KW Treponema pallidum.
XX WO9859034-A2.
PN 30-DEC-1998.
XX 23-JUN-1998; 98WO-US13041.
XX 24-JUN-1997; 97US-0050667.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Fraser CM;
XX WPI; 1999-081273/07.
XX New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX Claim 1; Page 402-408; 1150pp; English.
XX AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX Sequence 11516 BP; 2703 A; 3270 C; 2947 G; 2585 T; 11 other;
SQ

Query Match 30.1%; Score 26.2; DB 20; Length 11516;
Best Local Similarity 60.6%; Pred. No. 32;
Matches 43; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 12 TATTCATGGAGCTTTCGCCAGGAGCAGTCAACCCGCCAGAAAGAAATGTAC 71
Db 2299 TTTTCGGTGGTTCATTCAGAGGGGAAATCAACGATCCACGAAAGAAACAC 2240

Qy 72 TCCTGAAGTGG 82
Db 2239 TCCTTCTCTGG 2229

RESULT 12
AAK65476/c
ID AAK65476 standard; DNA; 27435 BP.
XX AAK65476;
AC AAK65476;
XX 06-NOV-2001 (first entry)
DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20288.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
PD 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR

PR 24-FEB-2000; 2000US-0184664.
PR* 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0218290.
PR 14-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

XX The present invention describes isolated anthranilate synthase alpha-
 CC subunit (ASAS), anthranilate synthase beta-subunit (ASBS), and
 CC tryptophan synthase alpha-subunit (TSAS) nucleic acids, and protein
 CC encoded by them, obtained from corn, rice, soybean and wheat cDNA
 CC libraries. The nucleic acid fragments may be used to create transgenic
 CC plants in which the disclosed ASAS, ASBS or TSAS are present at higher or
 CC lower levels than normal or in cell types or developmental stages in
 CC which they are not normally found. This would have the effect of altering
 CC the level of tryptophan in those cells. Manipulation of the levels of
 CC some of the ASASs will also results in changes in the response to
 CC pathogen attack. Because this pathway is not followed for the production
 CC of tryptophan in higher animals, these enzymes are very good candidates
 CC for the discovery of herbicides and fungicides. The ASAS, ASBS or TSAS
 CC can be used as targets to facilitate design and/or identification of
 CC inhibitors of those enzymes that may be useful as herbicides. Nucleic
 CC acid fragments can also be used as probes for genetically and physically
 CC mapping the genes that they are a part of, and as markers for traits
 CC linked to those genes. Such information may be useful in plant breeding
 CC in order to develop lines with desired phenotypes. AAZ25109 to AAZ25127
 CC represent specifically claimed nucleic acids from the present invention
 CC and AAZ42112 to AAZ42130 represent the proteins encoded by them.
 XX
 SQ Sequence 1133 BP; 279 A; 291 C; 308 G; 254 T; 1 other;

Query Match 29.9%; Score 26; DB 20; Length 1133;
 Best Local Similarity 62.1%; Pred. No. 17;
 Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 15 TCCAAATGGAGCCTTCGGCCAGGAGCGAGTCAACCCCGAGAGAAAGAAATGCTACTCC 74
 Db 757 TCCAGGTGTCCAAATTCACCCGAGAGCATCATCCCTGAAGGCAAGAAATCATCC 816

QY 75 TGAAGT 80
 Db 817 TCAACT 822

RESULT 14
 ABA43515
 ID ABA43515 standard; DNA; 473 BP.

XX ABA43515;
 XX
 XX
 DT 01-FEB-2002 (first entry)
 XX
 XX Human breast cell single exon nucleic acid probe #2210.
 XX
 XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer; ss.
 XX
 XX Homo sapiens.

XX
 XX WO200157271-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000662.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 FI WPI; 2001-496933/54.
 XX
 XX

PT New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 XX Claim 1; SEQ ID NO 2210; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BR 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 473 BP; 177 A; 95 C; 92 G; 109 T; 0 other;

Query Match 29.7%; Score 25.8; DB 22; Length 473;
 Best Local Similarity 58.4%; Pred. No. 15;
 Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCCAATGGGAGCCTTCGGCCAGGAGCGAGTCAACCCCGAGAAGA 60
 Db 18 GAACCTATTAAAGTAGCAATTCGAAATTTCTAGACCTAATTAAAGACACAGAAGT 77

QY 61 AAAGAATGTACTCCTGA 77
 Db 78 AAAGAAATTACTTCAGA 94

RESULT 15
 ABA53965
 ID ABA53965 standard; DNA; 473 BP.

XX ABA53965;
 XX
 XX 01-FEB-2002 (first entry)
 XX
 XX Human foetal liver single exon nucleic acid probe #2270.
 XX
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX
 XX Homo sapiens.
 XX
 XX WO200157277-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US000669.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 FI WPI; 2001-483447/52.
 XX
 XX

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 2270; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 473 BP; 177 A; 95 C; 92 G; 109 T; 0 other;

Query Match 29.7%; Score 25.8; DB 22; Length 473;
Best Local Similarity 58.4%; Pred. No. 15;
Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 GCAATATCATATTCGAATGGAGCCTTCGGCCAGGAGGTCACCCCGAGA 60
Db 18 GAACCTATTAAGTAGCAATTGCAGAAATTTGCTAGACGTAATTAAAGACACAAGAAGT 77

QY 61 AAAGAATGTACTCCTGA 77
Db 78 AAAGAATTAATTCTCAGA 94

Search completed: April 15, 2003, 18:58:42
Job time : 38.2159 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 ; Search time 197.114 Seconds
(without alignments)
12845.099 Million cell updates/sec

Title: US-09-647-019-8

Perfect score: 87

Sequence: 1 gcaaatatcaatattccaat.....gtactctgaaggaggag 87

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:

2: gb_htg:

3: gb_in:

4: gb_om:

5: gb_ov:

6: gb_pat:

7: gb_ph:

8: gb_pl:

9: gb_pr:

10: gb_ro:

11: gb_sta:

12: gb_sy:

13: gb_un:

14: gb_vi:

15: em_ba:

16: em_fun:

17: em_hum:

18: em_in:

19: em_mu:

20: em_om:

21: em_or:

22: em_ov:

23: em_pat:

24: em_ph:

25: em_pl:

26: em_ro:

27: em_sta:

28: em_un:

29: em_vi:

30: em_htg_hum:

31: em_htg_inv:

32: em_htg_other:

33: em_htg_mus:

34: em_htg_pln:

35: em_htg_rod:

36: em_htg_mam:

37: em_htg_vrt:

38: em_sy:

39: em_htgo_hum:

40: em_htgo_mus:

41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	87	100.0	835	9	BC005948	BC005948 Homo sapi
2	87	100.0	885	9	HS2250584	AJ250584 Homo sapi
3	87	100.0	886	6	AX322774	AX322774 Sequence
4	87	100.0	886	9	AF129505	AF129505 Homo sapi
5	87	100.0	47440	9	U73509	U73509 Homo sapien
6	87	100.0	150319	2	AL772370	AL772370 Homo sapi
7	66.2	76.1	787	10	AY026524	AY026524 Mus muscu
8	66.2	76.1	936	10	AF364070	AF364070 Mus muscu
9	66.2	76.1	943	10	MMU245772	AJ245772 Mus muscu
10	59.8	68.7	892	10	AF364071	AF364071 Rattus no
11	59.8	68.7	82586	2	AC096040	AC096040 Rattus no
12	59.8	68.7	188670	2	AC127606	AC127606 Rattus no
13	56.6	65.1	923	5	AF343894	AF343894 Xenopus l
14	36.4	41.8	50479	9	AC079455	AC079455 Homo sapi
15	36.4	41.8	108969	9	AC004813	AC004813 Homo sapi
16	36.4	41.8	179517	2	AC026361	AC026361 Homo sapi
17	31.6	36.3	164921	8	AF022186	AF022186 Cyanidium
18	31.4	36.1	117818	4	AC087180	AC087180 Sus scrof
19	31.2	35.9	155827	9	AC046136	AC046136 Homo sapi
20	31.2	35.9	157356	2	AC021376	AC021376 Homo sapi
21	30.2	34.7	178046	2	AC107182	AC107182 Rattus no
22	30.2	34.7	180148	2	AC131462	AC131462 Rattus no
23	29.8	34.3	120	9	HSADHSC10	AF026862 Homo sapi
24	29.8	34.3	332	11	G49400	G49400 sWSS4849 Hu
25	29.8	34.3	32173	9	AC004416	AC004416 Homo sapi
26	29.8	34.3	65201	9	AC004059	AC004059 Homo sapi
27	29.8	34.3	107609	9	AC118062	AC118062 Homo sapi
28	29.8	34.3	157559	2	AC021524	AC021524 Homo sapi
29	29.8	34.3	161673	2	AC129414	AC129414 Rattus no
30	29.8	34.3	194000	2	AC000016	AC000016 Homo sapi
31	29.8	34.3	202904	2	AP002077	AP002077 Homo sapi
32	29.8	34.3	203696	4	AC087860	AC087860 Bos tauru
33	29.6	34.0	54565	9	AC093585	AC093585 Homo sapi
34	29.4	33.8	171779	2	AC016000	AC016000 Homo sapi
35	29.4	33.8	177714	2	AC084790	AC084790 Homo sapi
36	29.4	33.8	182430	9	AL590381	AL590381 Human DNA
37	29.2	33.6	149632	2	AC120076	AC120076 Rattus no
38	28.8	33.1	55703	9	AL512448	AL512448 Human DNA
39	28.8	33.1	100987	2	AC112362	AC112362 Rattus no
40	28.8	33.1	190082	2	AC095758	AC095758 Rattus no
41	28.8	33.1	200395	2	AC125589	AC125589 Rattus no
42	28.6	32.9	85868	2	AC111392	AC111392 Rattus no
43	28.6	32.9	91323	9	AC106901	AC106901 Homo sapi
44	28.6	32.9	154519	2	AC107334	AC107334 Rattus no
45	28.4	32.6	273114	2	AC122820	AC122820 Mus muscu

ALIGNMENTS

RESULT 1

BC005948

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BC005948 Homo sapiens, small muscle protein, X-linked, clone MGC:14584
IMAGE:4246501, mRNA, complete cds.

BC005948

BC005948.1 GI:13543590

MGC.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 835)

Strausberg,R.

Direct Submission

835 bp mRNA linear PRI 12-JUL-2001
MGC:14584

JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-x@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 21 Row: a Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6625646.
Location/Qualifiers

FEATURES

source

1..835
/organism="Homo sapiens"
/db_xref="LocusID:23676"
/clone="MGC:14584 IMAGE:4246501"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
113..379
/codon_start=1
/product="small muscle protein, X-linked"
/protein_id="AAH05948.1"
/db_xref="GI:13543591"
/translation="MMSKQPVSNVRAIQANINIPMGAFPGAGOPRRKECTPEVEE
GVPTSDREKKPIPGAKLPGPAVNLSIQNIKSELKYVPKAEQ"

BASE COUNT 282 a 155 c 171 g 227 t

CDS

Query Match 100.0%; Score 87; DB 9; Length 835;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAAATATCAATATTCGAATGGAGCCTTCGCGCAGGAGCAGGTCAACCCCGAGAAGA 60
Db 158 GCAAAATATCAATATTCGAATGGAGCCTTCGCGCAGGAGCAGGTCAACCCCGAGAAGA 217

QY 61 AAGAATGTACTCTCGAGTGGAGGAG 87
Db 218 AAGAATGTACTCTCGAGTGGAGGAG 244

RESULT 2

HSA250584 885 bp mRNA linear PRI 12-APR-2001
LOCUS Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
DEFINITION protein (Srmx gene).
ACCESSION AJ250584
VERSION AJ250584.1 GI:10178976
KEYWORDS Srmx gene; stretch responsive muscle (X-chromosome).
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS Kemp T.J., Sadusky T.J., Simon, M., Brown, R., Eastwood, M., Sassoon, D.A. and Coulton, G.R.
TITLE Identification of a novel stretch-responsive skeletal muscle gene (Smpx)
JOURNAL Genomics 72 (3), 260-271 (2001)

JOURNAL Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial College of Science Technology and Medicine, SAF Bldg, Level 2, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
Related sequences: AJ245772, U73508 to U73509.

COMMENT

FEATURES

source

1..885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="skeletal muscle"
1..885
/gene="Srmx"
1..183
/gene="Srmx"
184..450
/gene="Srmx"
/codon_start=1
/product="stretch responsive muscle (X-chromosome)"
/protein_id="CAC08492.1"
/db_xref="GI:10178977"
/translation="MMSKQPVSNVRAIQANINIPMGAFPGAGOPRRKECTPEVEE
GVPTSDREKKPIPGAKLPGPAVNLSIQNIKSELKYVPKAEQ"
451..885
/gene="Srmx"
857..862
/gene="Srmx"

BASE COUNT 278 a 172 c 191 g 244 t

ORIGIN

Query Match 100.0%; Score 87; DB 9; Length 885;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAAATATCAATATTCGAATGGAGCCTTCGCGCAGGAGCAGGTCAACCCCGAGAAGA 60
Db 229 GCAAAATATCAATATTCGAATGGAGCCTTCGCGCAGGAGCAGGTCAACCCCGAGAAGA 288

QY 61 AAGAATGTACTCTCGAGTGGAGGAG 87
Db 289 AAGAATGTACTCTCGAGTGGAGGAG 315

RESULT 3

AX322774 886 bp DNA linear PAT 07-JAN-2002
LOCUS Sequence 18 from Patent WO0192567.
DEFINITION AX322774
ACCESSION AX322774
VERSION AX322774.1 GI:18093754
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1
AUTHORS Bunk, D., Reuner, B., Beck, J. and Henkel, T.
TITLE Novel target genes for diseases of the heart
JOURNAL Patent: WO 0192567-A 18 06-DEC-2001;
Medigene AG (DE)
FEATURES
source
1..886
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 278 a 172 c 191 g 245 t

ORIGIN

Query Match 100.0%; Score 87; DB 6; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.3e-20;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAAATATCAATATTCGAATGGAGCCTTCGCGCAGGAGCAGGTCAACCCCGAGAAGA 60

Db 229 GCAAAATATCAATATTCCTGAATGGAGCCCTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 288

Qy 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87

Db 289 AAAGAATGTACTCTCTGAAGTGGAGGAG 315

RESULT 4

AF129505 886 bp mRNA linear PRI 22-DEC-1999

LOCUS Homo sapiens small muscular protein (SMPX) mRNA, complete cds.

DEFINITION AF129505

ACCESSION AF129505.1 GI:6625646

VERSION

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 886)

TITLE Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.

JOURNAL Identification, mapping, and genomic structure of a novel

MEDLINE X-chromosomal human gene (SMPX) encoding a small muscular protein

20065879 Hum. Genet. 105 (5), 506-512 (1999)

REFERENCE 1059820

2 (bases 1 to 886)

AUTHORS Patzak,D.

TITLE Direct Submission

JOURNAL Submitted (18-JAN-1999) Molecular Human Genetics, Institut for

Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany

FEATURES

source 1..886

Gene /organism="Homo sapiens"

5'UTR /db xref="taxon:9606"

CDS /chromosome="X"

1..886 /map="Xq22.1-22.2"

5'UTR <1..>886

CDS /gene="SMPX"

1..183 /gene="SMPX"

184..450 /gene="SMPX"

misc_feature /codon_start=1

190 /evidence=not_experimental

3'UTR /product="small muscular protein"

polyA_signal /protein_id="AA19343.1"

BASE COUNT 279 a 172 c 191 g 244 t

ORIGIN

Query Match 100.0%; Score 87; DB 9; Length 886;

Best Local Similarity 100.0%; Pred. No. 1.3e-20;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAAAATATCAATATTCCTGAAGTGGAGCCCTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 60

Db 229 GCAAAATATCAATATTCCTGAAGTGGAGCCCTTCGGCCAGGAGCAGGTCAACCCCCCAGAAGA 288

Qy 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87

Db 289 AAAGAATGTACTCTCTGAAGTGGAGGAG 315

```
repeat_region /db_xref="taxon:9606"
repeat_region /chromosome="X"
repeat_region /map="Xp22.1-22.2"
repeat_region /clone="U228D4"
repeat_region /clone_lib="LLOXNCC01-U"
repeat_region 9..261
repeat_region /rpt_family="Retroviral"
repeat_region 2020..2180
repeat_region /rpt_family="MIR"
repeat_region 4824..4862
repeat_region /rpt_family="(CAAT)n"
repeat_region complement(5131..5222)
repeat_region /note="match to EST N87511 (NID:g1440713)"
repeat_region /note="match to EST W07478 (NID:g1281479) za96c10.rl"
repeat_region /note="match to EST W07478 (NID:g1281479) za96c10.rl"
repeat_region /note="match to EST AA092554 (NID:g1637327)"
repeat_region /note="match to EST AA248485 (NID:g1879506)"
repeat_region 7616..7763
repeat_region /rpt_family="MIR"
repeat_region 8781..8842
repeat_region /rpt_family="U2"
repeat_region 11721..11940
repeat_region /rpt_family="MER1_type"
repeat_region 12770..12961
repeat_region /rpt_family="MIR"
repeat_region 13039..13107
repeat_region /rpt_family="L2"
repeat_region 13897..14201
repeat_region /rpt_family="Alu"
repeat_region 14233..14397
repeat_region /rpt_family="MIR"
repeat_region complement(15622..15689)
repeat_region /note="match to EST AA248485 (NID:g1879506)"
repeat_region complement(15622..15689)
repeat_region /note="match to EST W07478 (NID:g1281479) za96c10.rl"
repeat_region 16751..16837
repeat_region /rpt_family="MIR"
repeat_region complement(19314..19497)
repeat_region /note="match to EST AA248485 (NID:g1879506)"
repeat_region complement(19314..19478)
repeat_region /note="match to EST W07478 (NID:g1281479) za96c10.rl"
repeat_region complement(19314..19447)
repeat_region /note="match to EST AA248067 (NID:g1878536)"
repeat_region complement(19314..19423)
repeat_region /note="match to EST N87511 (NID:g1440713)"
repeat_region complement(19314..19413)
repeat_region /note="match to EST AA092554 (NID:g1637327)"
repeat_region 21260..21572
repeat_region /rpt_family="Retroviral"
repeat_region 21984..22282
repeat_region /rpt_family="Alu"
repeat_region 23499..23673
repeat_region /rpt_family="MIR"
repeat_region 24236..24556
repeat_region /rpt_family="L1"
repeat_region 24759..24869
repeat_region /rpt_family="L1"
repeat_region 25920..26860
repeat_region /rpt_family="L1"
repeat_region 26867..27210
repeat_region /rpt_family="MaLR"
repeat_region 27213..27810
repeat_region /rpt_family="L1"
repeat_region 28727..28816
repeat_region /rpt_family="Retroviral"
repeat_region 28945..29005
repeat_region /rpt_family="Retroviral"
repeat_region 29715..30007
repeat_region /rpt_family="Alu"
repeat_region 30773..31151
repeat_region /rpt_family="MaLR"
```

```
repeat_region 31705..33575
repeat_region /rpt_family="L1"
repeat_region 33587..37330
repeat_region /rpt_family="L1"
repeat_region 37345..37464
repeat_region /rpt_family="(TA)n"
repeat_region 37504..37858
repeat_region /rpt_family="L1"
repeat_region 38102..38146
repeat_region /rpt_family="MIR"
repeat_region 38578..38706
repeat_region /rpt_family="L2"
repeat_region 38794..38914
repeat_region /rpt_family="Retroviral"
repeat_region 38939..39240
repeat_region /rpt_family="Alu"
repeat_region 39241..39431
repeat_region /rpt_family="Retroviral"
repeat_region 39469..39659
repeat_region /rpt_family="MER1_type"
repeat_region 39877..39946
repeat_region /rpt_family="L2"
repeat_region 40027..40328
repeat_region /rpt_family="Alu"
repeat_region 42307..42488
repeat_region /rpt_family="MER1_type"
repeat_region 43876..44349
repeat_region /rpt_family="MaLR"
repeat_region 45063..45245
repeat_region /rpt_family="MIR"
repeat_region 45276..45399
repeat_region /rpt_family="MER1_type"
repeat_region 46393..46517
repeat_region /rpt_family="L2"
repeat_region 46645..47185
repeat_region /rpt_family="MER21_g"
BASE COUNT 15150 a 9644 c 9310 g 13336 t
ORIGIN
Query Match 100.0%; Score 87; DB 9; Length 47440;
Best Local Similarity 100.0%; Pred. No. 2.3e-20;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAATATCAATATTCATATGGAGCCTTTGGCCAGGAGCAGGTCAACCCCCAGAGA 60
Db 5220 GCAATATCAATATTCATATGGAGCCTTTGGCCAGGAGCAGGTCAACCCCCAGAGA 5161
QY 61 AAGAATGTACTCTCTGAAGTGGAGGAG 87
Db 5160 AAGAATGTACTCTCTGAAGTGGAGGAG 5134
RESULT 6
AL772370/c 150319 bp DNA linear HTG 17-AUG-2002
LOCUS Homo sapiens chromosome X clone RP11-184B10, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 unordered pieces.
ACCESSION AL772370
VERSION AL772370.5 GI:22416024
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150319)
AUTHORS Bird,C.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:2204612.
----- Genome Center
Center: Wellcome Trust Sanger Institute
```

Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: bal18B10
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 99% of reads
 quality: 149793 bases at least Q40
 Consensus quality: 149914 bases at least Q30
 Consensus quality: 150006 bases at least Q20
 Insert size: 150219; sum-of-contigs
 quality coverage: 17.18x in Q20 bases; agarose-fp
 coverage: 17.58x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 4859: contig of 4859 bp in length
 * 4860 4959: gap of 100 bp
 * 4960 150319: contig of 145360 bp in length.

FEATURES

source

1..150319
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP11-184B10"
 /clone_lib="RPC1-11.1"

misc_feature

1..4859
 /note="assembly fragment:05115"

misc_feature

4960..150319
 /note="assembly fragment:05270"
 /fragment_chain:1
 clone end:SP6
 vector_side:right"

BASE COUNT 47436 a 29741 c 29722 g 43320 t 100 others

Query Match

Best Local Similarity 100.0%; Score 87; DB 2; Length 150319;
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGAGCCTTTCCGCGAGGAGCAGGTCAACCCCGAGAGA 60
 Db 17993 GCAATATCAATATTCATGGAGCCTTTCCGCGAGGAGCAGGTCAACCCCGAGAGA 17934

QY 61 AAGAATGTACTCTCGAGTGGAGGAG 87 ✓
 Db 17933 AAGAATGTACTCTCGAGTGGAGGAG 17907

RESULT 7

AY026524

LOCUS 787 bp mRNA linear ROD 28-JUN-2001
 DEFINITION Mus musculus muscle-specific protein CSL (Cel) mRNA, complete cds.
 ACCESSION AY026524
 VERSION AY026524.1 GI:14575061

KEYWORDS

SOURCE Mus musculus.

ORGANISM

Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 787)

REFERENCE

AUTHORS Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C.,
 Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F.,

TITLE

The small muscle-specific protein Cel modifies cell shape and
 promotes myocyte fusion in an insulin-like growth factor
 1-dependent manner

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Mohun,T. and Harvey,R.P.

The small muscle-specific protein Cel modifies cell shape and
 promotes myocyte fusion in an insulin-like growth factor
 1-dependent manner

J. Cell Biol. 153 (5), 985-998 (2001)

21275706

11381084

2 (bases 1 to 787)

Palmer,S., Groves,N., Schindeler,A., Yeoh,T., Biben,C., Wang,C.-C.,

Sparrow,D.B., Barnett,L., Jenkins,N.A., Copeland,N.G., Koentgen,F.,

Mohun,T. and Harvey,R.P.

Direct Submission

Submitted (01-FEB-2001) Developmental Biology, Victor Chang Cardiac

Research Institute, 384 Victoria St, Darlinghurst, Sydney, New

South Wales 2010, Australia

Location/Qualifiers

1..787

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/chromosome="X"

1..787

/gene="Cel"

206..463

/gene="Cel"

/note="ortholog of Homo sapiens SMPX"

/codon_start=1

/product="muscle-specific protein CSL"

/protein_id="AAK07682.1"

/db_xref="GI:14575062"

/translation="MSKQPIISNRAIQININPMGAPRPGAGQPPRRKESTPTEEGA

PTTSEKAPDIPGMKPGVFNLSIQNKSELKVPKGEQ"

BASE COUNT 233 a 167 c 185 g 202 t

ORIGIN

Query Match 76.1%; Score 66.2; DB 10; Length 787;

Best Local Similarity 85.1%; Pred. No. 4.8e-13;

Matches 74; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATGGAGCCTTTCCGCGAGGAGCAGGTCAACCCCGAGAGA 60

Db 245 GCGAATATCAATATTCATGGAGCCTTTCCGCGAGGAGCAGGTCAACCCCGAGAGA 304

QY 61 AAGAATGTACTCTCGAGTGGAGGAG 87

Db 305 AAGAATGTACTCTCGAGTGGAGGAG 331

RESULT 8

AF364070

LOCUS 936 bp mRNA linear ROD 04-MAY-2001

DEFINITION Mus musculus SMPX protein (Smpx) mRNA, complete cds.

ACCESSION AF364070

VERSION AF364070.1 GI:13940507

KEYWORDS

SOURCE Mus musculus.

Mus musculus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 936)

Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.

Identification, mapping, and genomic structure of a novel

X-chromosomal human gene (SMPX) encoding a small muscular protein

Hum. Genet. 105 (5), 506-512 (1999)

20065879

10598820

2 (bases 1 to 936)

Patzak,D.

Direct Submission

Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inet. f. Humang.,

Fleischmannstr. 42-44, D-17487 Greifswald, Germany

Location/Qualifiers

1..936

```
/organism="Mus musculus"
/db_xref="taxon:10090"
<1..936
/gene="Smpx"
199..456
/gene="Smpx"
/codon_start=1
/product="SMX protein"
/protein_id="AAK50398.1"
/db_xref="GI:13940508"
/translations="MSKOPTSNVRAIQANINIPMGAFRPGAGQPPRRKESTPTEBEGA
PTTSEKKPIPGMKKFPGPVNLSEIQNVKSELKFPVKGEQ"
polyA_signal
906..911
/gene="Smpx"
/evidence=experimental
BASE COUNT 283 a 187 c 203 g 263 t
ORIGIN
Query Match 76.1%; Score 66.2; DB 10; Length 936;
Best Local Similarity 85.1%; Pred. No. 4.9e-13;
Matches 74; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1 GCAATATCAATATTCGAATGGAGCGCTTCGGCAGGAGCAGGTCAACCCCGCAGAAGA 60
Db 238 CGAATATCAATATTCGAATGGAGCGCTTCGGCAGGAGTGGCGAGCTCCCGAGG 297
QY 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87
Db 298 AAAGAGAGTACTCTCTGAAGTGGAGGAG 324
RESULT 9
MMU245772 943 bp mRNA linear ROD 12-APR-2001
LOCUS Mus musculus mRNA for stretch responsive muscle (X-chromosome)
DEFINITION protein (Smpx gene).
ACCESSION AJ245772
VERSION AJ245772.1 GI:10178962
KEYWORDS Smpx gene; stretch responsive muscle (X-chromosome).
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 943)
Kemp, T.J., Sadusky, T.J., Simon, M., Brown, R., Eastwood, M.,
Sassoon, D.A. and Coulton, G.R.
Identification of a novel stretch-responsive skeletal muscle gene
(Smpx)
Genomics 72 (3), 260-271 (2001)
JOURNAL 11401441
MEDLINE 21295047
PUBMED 11401441
REFERENCE 2 (bases 1 to 943)
AUTHORS Kemp, T.J.
TITLE Direct Submission
JOURNAL Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial
College School of Medicine, SAF Building, Exhibition Road, South
Kensington, London SW7 2AZ, UNITED KINGDOM
FEATURES
Location/Qualifiers
1..943
/organism="Mus musculus"
/strain="C57BL/10"
/db_xref="taxon:10090"
/tissue_type="skeletal muscle"
1..943
/gene="Smpx"
1..214
/gene="Smpx"
215..472
/gene="Smpx"
/codon_start=1
/product="stretch responsive muscle (X-chromosome)"
/protein_id="CAC08493.1"
/db_xref="GI:10178963"
```

```
/translation="MSKOPTSNVRAIQANINIPMGAFRPGAGQPPRRKESTPTEBEGA
PTTSEKKPIPGMKKFPGPVNLSEIQNVKSELKFPVKGEQ"
473..943
/gene="Srmx"
535..539
/gene="Srmx"
/function="mRNA destabilising motif"
585..590
/gene="Srmx"
/function="mRNA destabilising motif"
658..662
/gene="Srmx"
/function="mRNA destabilising motif"
706..710
/gene="Srmx"
/function="mRNA destabilising motif"
740..744
/gene="Srmx"
/function="mRNA destabilising motif"
779..785
/gene="Srmx"
/function="mRNA destabilising motif"
909..913
/gene="Srmx"
/function="mRNA destabilising motif"
923..928
/gene="Srmx"
polyA_signal
943
polyA_site
/gene="Srmx"
BASE COUNT 278 a 192 c 210 g 262 t 1 others
ORIGIN
Query Match 76.1%; Score 66.2; DB 10; Length 943;
Best Local Similarity 85.1%; Pred. No. 4.9e-13;
Matches 74; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1 GCAATATCAATATTCGAATGGAGCGCTTCGGCAGGAGCAGGTCAACCCCGCAGAAGA 60
Db 254 CGAATATCAATATTCGAATGGAGCGCTTCGGCAGGAGTGGCGAGCTCCCGAGG 313
QY 61 AAAGAATGTACTCTCTGAAGTGGAGGAG 87
Db 314 AAAGAGAGTACTCTCTGAAGTGGAGGAG 340
RESULT 10
AF364071 892 bp mRNA linear ROD 04-MAY-2001
LOCUS Rattus norvegicus SMPX protein (Smpx) mRNA, complete cds.
DEFINITION AF364071
ACCESSION AF364071
VERSION AF364071.1 GI:13940509
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 892)
Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.
Identification, mapping, and genomic structure of a novel
X-chromosomal human gene (SMPX) encoding a small muscular protein
Hum. Genet. 105 (5), 506-512 (1999)
JOURNAL 20065879
MEDLINE 10598820
PUBMED 10598820
REFERENCE 2 (bases 1 to 892)
AUTHORS Patzak, D.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
Fleischmannstr. 42-44, D-17487 Greifswald, Germany
FEATURES
Location/Qualifiers
1..892
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
```

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenwo, S., Oguh, M., Okwuonu, G., Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sleson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.P., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 82586)
Worley, K.C.

Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 82586)
Worley, K.C.

Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17943701.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GBRG
Center clone name: CH230-291f
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.903029
Consensus quality: 35907 bases at least Q40
Consensus quality: 40324 bases at least Q30
Consensus quality: 44113 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1318 1417: contig of 1317 bp in length
1418 2800: contig of 1383 bp in length
2801 2900: gap of unknown length
2901 4244: contig of 1344 bp in length
4245 4345: gap of unknown length
4345 5864: contig of 1520 bp in length
5865 5965: gap of unknown length
5965 6968: contig of 1004 bp in length
6969 7069: gap of unknown length
7069 8167: contig of 1098 bp in length
8167 8267: gap of unknown length
8267 9384: contig of 1118 bp in length
9385 9485: gap of unknown length
9485 10630: contig of 1146 bp in length
10631 10730: gap of unknown length
10731 11858: contig of 1128 bp in length
11859 11859: gap of unknown length

```
* 11959 13413: contig of 1455 bp in length
* 13414 13513: gap of unknown length
* 14847 14847: contig of 1334 bp in length
* 14848 14947: gap of unknown length
* 16180 16180: contig of 1233 bp in length
* 16280 16280: gap of unknown length
* 16281 17804: contig of 1524 bp in length
* 17805 17904: gap of unknown length
* 17905 18913: contig of 1009 bp in length
* 18914 19013: gap of unknown length
* 20050 20050: contig of 1037 bp in length
* 20150 20150: gap of unknown length
* 21378 21378: contig of 1228 bp in length
* 21478 21478: gap of unknown length
* 22828 22828: contig of 1350 bp in length
* 22928 22928: gap of unknown length
* 23998 23998: contig of 1070 bp in length
* 24098 24098: gap of unknown length
* 25598 25598: contig of 1498 bp in length
* 25698 25698: gap of unknown length
* 26965 26965: contig of 1269 bp in length
* 27065 27065: gap of unknown length
* 29291 29291: contig of 2226 bp in length
* 29391 29391: gap of unknown length
* 30870 30870: contig of 1479 bp in length
* 30970 30970: gap of unknown length
* 33590 33590: contig of 2620 bp in length
* 33690 33690: gap of unknown length
* 34829 34829: contig of 1139 bp in length
* 34929 34929: gap of unknown length
* 36600 36600: contig of 1671 bp in length
* 36700 36700: gap of unknown length
* 38157 38157: contig of 1457 bp in length
* 38257 38257: gap of unknown length
* 39358 39358: contig of 1101 bp in length
* 39458 39458: gap of unknown length
* 41104 41104: contig of 1646 bp in length
* 42104 42104: gap of unknown length
* 42418 42418: contig of 1214 bp in length
* 42518 42518: gap of unknown length
* 43862 43862: contig of 1344 bp in length
* 43962 43962: gap of unknown length
* 45961 45961: contig of 1999 bp in length
* 46061 46061: gap of unknown length
* 48791 48791: contig of 2730 bp in length
* 48891 48891: gap of unknown length
* 51571 51571: contig of 2680 bp in length
* 51671 51671: gap of unknown length
* 52660 52660: gap of unknown length
* 53161 53161: contig of 1453 bp in length
* 54713 54713: gap of unknown length
* 54813 54813: gap of unknown length
* 56886 56886: contig of 1873 bp in length
* 56786 56786: gap of unknown length
* 58525 58525: contig of 1739 bp in length
* 58625 58625: gap of unknown length
* 61362 61362: contig of 2737 bp in length
* 61462 61462: gap of unknown length
* 64676 64676: contig of 3214 bp in length
* 64776 64776: gap of unknown length
* 67299 67299: contig of 2523 bp in length
* 67399 67399: gap of unknown length
* 69326 69326: contig of 1927 bp in length
* 69426 69426: gap of unknown length
* 72187 72187: contig of 2761 bp in length
* 72287 72287: gap of unknown length
* 73531 73531: contig of 1244 bp in length
* 73631 73631: gap of unknown length
* 76752 76752: contig of 3121 bp in length
* 76852 76852: gap of unknown length
* 79280 79280: contig of 2428 bp in length
* 79380 79380: gap of unknown length
* 82586 82586: contig of 3206 bp in length.

FEATURES             source
Location/Qualifiers
1..82586
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-2911"
BASE COUNT  23112 a 17332 c 15913 g 21644 t  4585 others
ORIGIN
Query Match      68.7%; Score 59.8; DB 2; Length 82586;
Best Local Similarity 80.5%; Pred. No. 1.9e-10;
Matches 70; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 1 GCAATATCAATATTCATGAGGCTTCGCGCAGGAGCAGGTCACACCCCGAGAGA 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32064 GCAATATTAATATTCATGAGGCTTCGCGGAGCTTCGCGGAGCTCCGAGAGG 32123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AAAGATGACTCTCTGGAAGTGGAGGAG 87
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 32124 AAAGAGATACCCCTGGAAGTGGAGGAG 32150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AC127606
LOCUS          188670 bp      DNA      linear      HTG 18-JUL-2002
DEFINITION    Rattus norvegicus clone CH230-20015, *** SEQUENCING IN PROGRESS
ACCESSION     AC127606
VERSION       AC127606.1 GI:21902814
KEYWORDS      HTG; HTGS PHASE1.
SOURCE        Norway rat.
ORGANISM      Rattus norvegicus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
               Rattus.
REFERENCE     1 (bases 1 to 188670)
AUTHORS       Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
               Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
               Barbara,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
               Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
               Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
               Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
               Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
               Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
               Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
               Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
               Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
               Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
               Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
               Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
               Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
               Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
               Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
               Homs,F., Howard,S., Huber,U., Hulyk,S., Hume,J., Jackson,L.E.,
               Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
               Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
               Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
               Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
               Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
               Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
               Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
               Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
               Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
               Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okwuonu,G.,
               Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
               Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Reny,G.,
               Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
               Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
               Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
               Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
               Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
               Umehani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
               Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
               Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,
```

Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 188670)
Worley, K.C.

Direct Submission

Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GXOD

Center clone name: CH230-20D15

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142614 bases at least Q40

Consensus quality: 149053 bases at least Q30

Consensus quality: 154556 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 60 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1

* 1104: contig of 1104 bp in length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 1205: contig of unknown length

* 30906: gap of unknown length

* 30906: contig of 1320 bp in length

* 32225: gap of unknown length

* 32225: contig of 1982 bp in length

* 34307: gap of unknown length

* 34307: contig of 1535 bp in length

* 35942: gap of unknown length

* 35942: contig of 2075 bp in length

* 36042: gap of unknown length

* 36042: contig of 1985 bp in length

* 38117: gap of unknown length

* 38117: contig of 2222 bp in length

* 40302: gap of unknown length

* 40302: contig of 1896 bp in length

* 42524: gap of unknown length

* 42524: contig of 1896 bp in length

* 44520: gap of unknown length

* 44520: contig of 3091 bp in length

* 47811: gap of unknown length

* 47811: contig of 2351 bp in length

* 50162: gap of unknown length

* 50162: contig of 1981 bp in length

* 52243: gap of unknown length

* 52243: contig of 2805 bp in length

* 55248: gap of unknown length

* 55248: contig of 2731 bp in length

* 57979: gap of unknown length

* 57979: contig of 2318 bp in length

* 60397: gap of unknown length

* 60397: contig of 2523 bp in length

* 63120: gap of unknown length

* 63120: contig of 3893 bp in length

* 67013: gap of unknown length

* 67013: contig of 1957 bp in length

* 69070: gap of unknown length

* 69070: contig of 2226 bp in length

* 71396: gap of unknown length

* 71396: contig of 2093 bp in length

* 73589: gap of unknown length

* 73589: contig of 2949 bp in length

* 76638: gap of unknown length

* 76638: contig of 2877 bp in length

* 79715: gap of unknown length

* 79715: contig of 3028 bp in length

* 82743: gap of unknown length

* 82743: contig of 3421 bp in length

* 86264: gap of unknown length

* 86264: contig of 3041 bp in length

* 89505: gap of unknown length

* 89505: contig of 2840 bp in length

* 92345: gap of unknown length

* 92345: contig of 2995 bp in length

* 95440: gap of unknown length

* 95440: contig of 4383 bp in length

* 99223: gap of unknown length

* 99223: contig of 4293 bp in length

* 100024: gap of unknown length

* 100024: contig of 3696 bp in length

* 104317: gap of unknown length

* 104317: contig of 3371 bp in length

* 108212: gap of unknown length

* 108212: contig of 4908 bp in length

* 111583: gap of unknown length

* 111583: contig of 3916 bp in length

* 116591: gap of unknown length

* 116591: contig of 5619 bp in length

* 120607: gap of unknown length

* 120607: contig of 4377 bp in length

* 126326: gap of unknown length

* 126326: contig of 6115 bp in length

* 130803: gap of unknown length

* 130803: contig of 5814 bp in length

* 137018: gap of unknown length

* 137018: contig of 6343 bp in length

* 142932: gap of unknown length

* 142932: contig of 6343 bp in length

* 149375: gap of unknown length

* 149375: contig of 6343 bp in length

```

* 149476 156271: contig of 6796 bp in length
* 156272 156371: gap of unknown length
* 156372 164838: contig of 8467 bp in length
* 164839 164938: gap of unknown length

Query Match
Best Local Similarity 68.7%; Score 59.8; DB 2; Length 188670;
Matches 70; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATATGGAGCCTTTCGGCCAGGAGCAGTCAACCCCGAGAAGA 60
Db 115912 GCCAATATTAATATTCATATGGAGCCTTTCGGCCAGGAGTGGCAGCTCCCGAAGG 115971

QY 61 AAAGAATGTACTCTCGAAGTGGAGGAG 87
Db 115972 AAAGAGATACCCCTGGAAGTGGAG 115998

RESULT 13
LOCUS AF343894 923 bp mRNA linear VRT 05-JUL-2001
DEFINITION Xenopus laevis Chisel (Csl) mRNA, complete cds.
ACCESSION AF343894
VERSION AF343894.1 GI:14599748
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 923)
Palmer, S., Groves, N., Schindeler, A., Yeoh, T., Biben, C., Wang, C.-C.,
Sparrow, D.B., Barnett, L., Jenkins, N.A., Copeland, N.G., Koentgen, F.,
Mohn, T.J. and Harvey, R.P.
The small muscle-specific protein Csl modifies cell shape and
promotes myocyte fusion in an insulin-like growth factor
1-dependent manner
J. Cell Biol. 153 (5), 985-998 (2001)
21275706
11381084
2 (bases 1 to 923)
Sparrow, D.B., Mohn, T.J. and Harvey, R.P.
Direct Submission
Submitted (30-JAN-2001) Developmental Biology Unit, Victor Chang
Cardiac Research Institute, 384 Victoria St, Darlinghurst, NSW
2010, Australia
FEATURES
source
1..923
/organism="Xenopus laevis"
/db_xref="taxon:8355"
1..923
/gene="Csl"
435..710
/gene="Csl"
/codon_start=1
/product="Chisel"
/protein_id="AAK71068.1"
/db_xref="GI:14599748"
/translation="MSQPASNIRSIQNIINPMGAPRPGAGQPKPKFSTEEHQV
PTPESEKSEKPIPGAVKLPGPAPNLSEIQNVKSLKFPKAEQ"
BASE COUNT 312 a 184 c 187 g 240 t
ORIGIN

Query Match
Best Local Similarity 65.1%; Score 56.6; DB 5; Length 923;
Matches 68; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 GCAATATCAATATTCATATGGAGCCTTTCGGCCAGGAGCAGTCAACCCCGAGAAGA 60
Db 474 GCTNATATCAATATTCATATGGAGCCTTTCGGCCAGGAGTGGCAGCTCCCGAAGG 533

QY 61 AAAGAATGTACTCTCGAAGTGGAGGAG 87

```

```

Db 534 AAAGATTTTCAACTCAAGAGCAGCAG 560
AC079455/c
RESULT 14
LOCUS AC079455 50479 bp DNA linear PRI 26-JUN-2002
DEFINITION Homo sapiens 12 BAC RP11-147D16 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC079455
VERSION AC079455.19 GI:21591798
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 50479)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chiu, Z., Chiu, D., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douchwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,
Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,
Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,
Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M.,
Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S.,
Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, F.,
Johnson, R., Jolivet, S., Joudah, S., Karlson, E., Kelly, S., Khan, U.,
King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N.,
Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O.,
Lieu, C., Liu, J., Liu, W., Loulsegod, H., Lozado, R.J., Lu, X.,
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,
Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,
Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S.,
Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,
Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M.,
Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokoko, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,
Scherer, S., Scott, G., Shen, H., Shim, C., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Zorrilla, S., Kucherlapati, R.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 50479)
Worley, K.C.
Direct Submission
Submitted (02-SEP-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 50479)
Worley, K.C.
Direct Submission
Submitted (07-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 50479)

```


AUTHORS

TITLE

JOURNAL

Worley, K.C.

Direct Submission

Submitted (26-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jun 26, 2002 this sequence version replaced gi:21327310.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annot> ation.html.

FEATURES

source

Location/Qualifiers

1..50479
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12"

88..177
 /clone="RP11-147D16"

178..317
 /rpt_family="ORSL"

324..612
 /rpt_family="FLAM_C"

613..714
 /rpt_family="AluSx"

715..1006
 /rpt_family="ORSL"

complement(1008..1075)
 /rpt_family="MER5A"

complement(1076..1333)
 /rpt_family="AluJb"

complement(1334..1408)
 /rpt_family="MER5A"

complement(1763..2073)
 /rpt_family="AluJb"

2080..2100
 /rpt_family="AluJb"

complement(2102..2346)
 /rpt_family="AluSg1"

2525..2826
 /rpt_family="AluSp"

2548..2754

repeat_region /standard_name="RH80030"
 3150..3290
 /rpt_family="MER5A"
 repeat_region 3292..3392
 /rpt_family="MER5B"
 repeat_region complement(3393..3672)
 /rpt_family="AluJb"
 repeat_region 3673..3684
 /rpt_family="MER5B"
 repeat_region complement(3685..3990)
 /rpt_family="AluSg"
 repeat_region 3991..4055
 /rpt_family="MER5B"
 repeat_region complement(4371..4498)
 /rpt_family="MIR"
 repeat_region 4768..5077
 /rpt_family="AluSx"
 repeat_region 5150..5323
 /rpt_family="L1MC/D"
 repeat_region complement(5443..5749)
 /rpt_family="AluSx"
 repeat_region complement(5825..5868)
 /rpt_family="MIR"
 repeat_region 5986..6116
 /rpt_family="AluJo"
 repeat_region 6356..6438
 /rpt_family="AT rich"
 repeat_region complement(7008..7139)
 /rpt_family="L2"
 repeat_region complement(7141..7217)
 /rpt_family="MLT1A1"
 repeat_region 7219..7271
 /rpt_family="(TCCCC)n"
 repeat_region complement(7272..7604)
 /rpt_family="AluSx"
 repeat_region complement(7607..7636)
 /rpt_family="MLT1A1"
 misc_feature 7608..7677
 /function="pcr product sequence only"
 repeat_region 7637..7939
 /rpt_family="AluSx"
 misc_feature 7640..7641
 /function="low quality"
 misc_feature 7643..7645
 /function="low quality"
 misc_feature 7656..7659
 /function="low quality"
 misc_feature 7669..7673
 /function="low quality"
 misc_feature 7699..7701
 /function="low quality"
 misc_feature 7742
 /function="low quality"
 misc_feature 7744..7749
 /function="low quality"
 misc_feature 7751..7752
 /function="low quality"
 misc_feature 7757..7758
 /function="low quality"

Query Match 41.8%; Score 36.4; DB 9; Length 50479;
 Best Local Similarity 70.0%; Pred. No. 0.057;

Matches 49; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 17 CAATGGAGCCTTTCCGCCAGGAGCAGTCAACCCCGAGAGAAAGAAATGACTCTG 76
 Db 14555 CAATGTAGCCATGAGCCAGTATCAGTTCAAGCAGCAGAGATAGATCCACTCTT 14496

Qy 77 AAGTGGAGGA 86

Db 14495 GAAGGGAGGA 14486

RESULT 15
AC004813 108969 bp DNA linear PRI 29-MAY-1999
LOCUS Homo sapiens clone 277F10, complete sequence.
DEFINITION AC004813
ACCESSION AC004813.2 GI:4926912
VERSION AC004813.2
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 108969)
Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 108969)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On May 29, 1999 this sequence version replaced gi:3213184.
FEATURES
source
1..108969
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="277F10"
BASE COUNT 32101 a 24676 c 23912 g 28280 t
ORIGIN
Query Match 41.8%; Score 36.4; DB 9; Length 108969;
Best Local Similarity 70.0%; Pred.No. 0.063;
Matches 49; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Qy 17 CAATGGGAGCCTTTCCGCCGAGGAGCGTCAACCCGCCGAGGAGAAATGTACTCTG 76
Db 95223 CAATGTGAGCCATGAGGCCAGTATCAGTTCAGCGGAGCAGAGAAATAGATCCACCTTT 95282
Qy 77 AAGTCGAGGA 86
Db 95283 GAAGGGAGGA 95292

Search completed: April 15, 2003, 20:57:25
Job time : 332.114 secs

AA211521
LOCUS AA211521 501 bp mRNA linear EST 31-JAN-1997
DEFINITION zn55b01.r1 Stragogene muscle 937209 Homo sapiens cDNA clone
IMAGE:562057 5', mRNA sequence.
ACCESSION AA211521
VERSION AA211521.1 GI:1810175
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, W., Hawkins
, M., Hiltman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore
, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Travaekis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Warr, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the

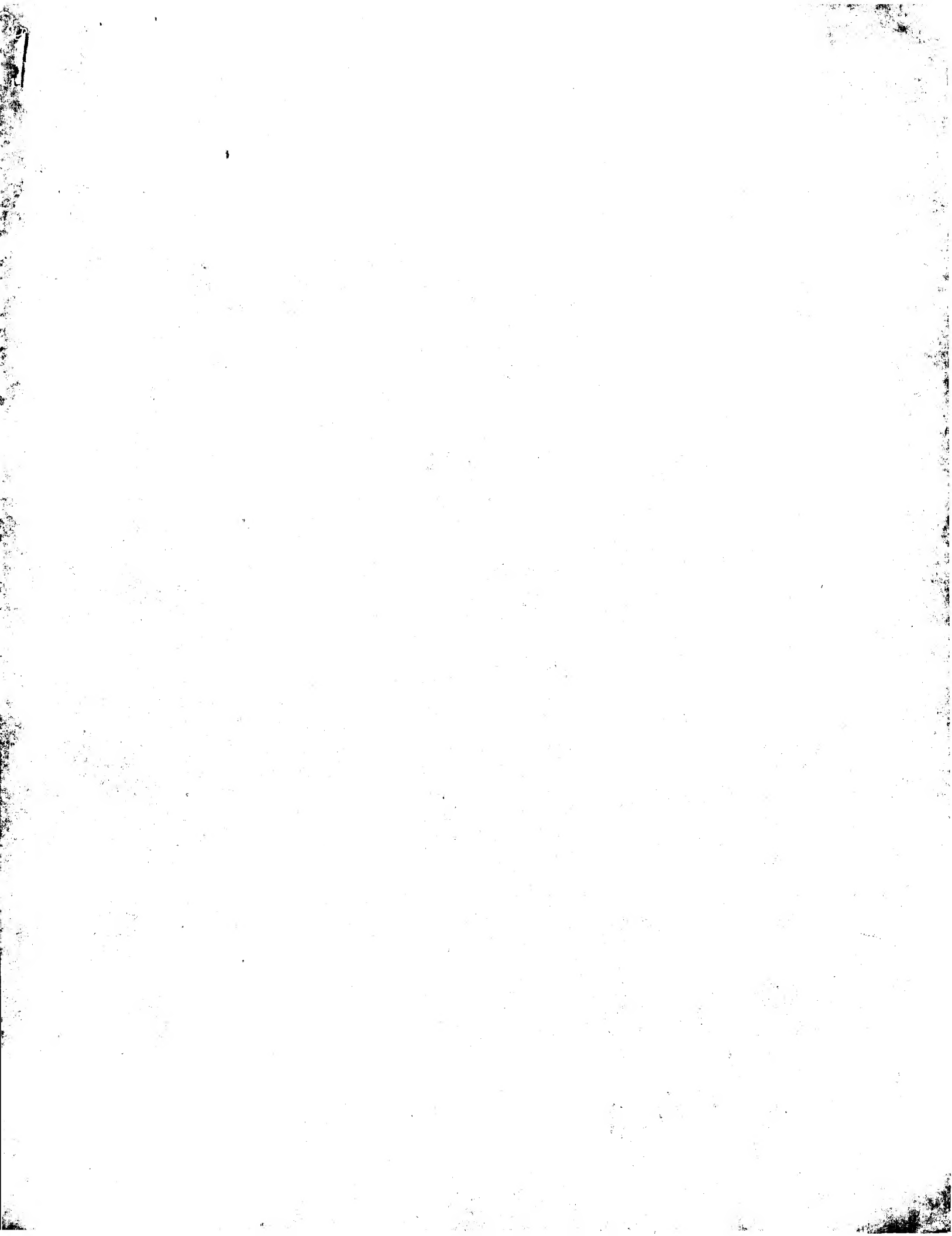


IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28ml3 rev1 ET from Amersham
 High quality sequence stop: 443.

FEATURES

SOURCE

Location/Qualifiers

1. 501

/organism="Homo sapiens"

/db_xref="GDB:4595347"

/db_xref="taxon:9606"

/clone="IMAGE:562057"

/clone_lib="Stratagene muscle 937209"

/ribose_type="muscle"

/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skeletal muscle; Vector: pBluescript SK-"

Site 1: EcoRI, Site 2: XhoI; Cloned unidirectionally.

Primer: Oligo dT, Skeletal muscle from patient with

malignant hyperthermia. Average insert size: 1.0 kb;

Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAG

3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

BASE COUNT 167 a 103 c 116 g 115 t

Query Match 100.0%; Score 149; DB 9; Length 501;
 Best Local Similarity 100.0%; Pred. No. 4.7e-31;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTGTTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 60
 |||

DB 229 GGTGTTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 288
 |||

OY 61 GGACCTGCAGTCAATCTATCGAAATCCAGATATTTAAATGAATTAATATGTCCCC 120
 |||

DB 289 GGACCTGCAGTCAATCTATCGAAATCCAGATATTTAAATGAATTAATATGTCCCC 348
 |||

OY 121 AAAGCTGAACAGTAGTAGAGAGAGAGAG 149
 |||

DB 349 AAAGCTGAACAGTAGTAGAGAGAGAGAG 377
 |||

RESULT 2 683 bp mRNA linear EST 29-SEP-2000
 LOCUS BE856875/c

DEFINITION BE856875.1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:3300004 3' similar to TR:Q9UHP9 Q9UHP9 SMALL MUSCULAR

PROTEIN: ; mRNA sequence.

ACCESSION BE856875.1 GI:10370341

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

1. 683

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3300004"

/clone_lib="Soares NSF_F8_9W_OT_PA_P_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHF pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HB-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326653 Soares NBHOT
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT

198 a 142 c 123 g 220 t

ORIGIN

Query Match 100.0%; Score 149; DB 12; Length 683;
 Best Local Similarity 100.0%; Pred. No. 4.7e-31;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGTGTTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 60
 |||

DB 567 GGTGTTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 508
 |||

OY 61 GGACCTGCAGTCAATCTATCGAAATCCAGATATTTAAATGAATTAATATGTCCCC 120
 |||

DB 507 GGACCTGCAGTCAATCTATCGAAATCCAGATATTTAAATGAATTAATATGTCCCC 448
 |||

OY 121 AAAGCTGAACAGTAGTAGAGAGAGAGAG 149
 |||

DB 447 AAAGCTGAACAGTAGTAGAGAGAGAGAG 419
 |||

RESULT 3

BM697544 719 bp mRNA linear EST 28-FEB-2002
 LOCUS UI-E-DX0-agn-1-12-0-UI_r1 UI-E-DX0 Homo sapiens cDNA clone

DEFINITION UI-E-DX0-agn-1-12-0-UI_r1 mRNA sequence.

ACCESSION BM697544 GI:19010802

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

SOURCE

1. 719

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-DX0-agn-1-12-0-UI"

/clone_lib="UI-E-DX0"

/ribose_type="fetal eyes"

/dev_stage="fetal"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-E-DX0 is a cDNA library containing the following
 tissue(s): fetal eyes. The library was constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tag for this library is AGAATCAGA. This library
 was created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI)."

BASE COUNT

226 a 142 c 163 g 186 t 2 others

ORIGIN

Query Match 100.0%; Score 149; DB 14; Length 719;
 Best Local Similarity 100.0%; Pred. No. 4.7e-31;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGTTCTCCGACCTCGATGAGGAGAAGCCATTCCAGAGCGAAGAACTTCCA 60
 DB 272 GGTGTTCTCCGACCTCGATGAGGAGAAGCCATTCCAGAGCGAAGAACTTCCA 60
 QY 61 GGACCTGAGTCAATCTATCGAAATTCAGAAATTAAGTGAATTAATATATGTCGCC 120
 DB 332 GGACCTGAGTCAATCTATCGAAATTCAGAAATTAAGTGAATTAATATATGTCGCC 120
 QY 121 AAAGCTGAACAGTAGTAGAAGAAAAAG 149
 DB 392 AAAGCTGAACAGTAGTAGAAGAAAAAG 149

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:52:31 ; Search time 309.035 Seconds

(without alignments)
7808.593 Million cell updates/sec

Title: US-09-647-019-9

Perfect score: 149
Sequence: 1 ggtgtctctccacacccgcga.....cagtagtaggaagaaaaag 149

Scoring table:

IDENTITY NUC
Gapex 10'-0, Gapex 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_gstba:*
2: em_gsthum:*
3: em_gstlin:*
4: em_gstmu:*
5: em_gstov:*
6: em_gstpl:*
7: em_gstto:*
8: em_gstc:*
9: gb_gstl:*
10: gb_gstc2:*
11: gb_gstc3:*
12: gb_gstc4:*
13: gb_gstc5:*
14: gb_gstc6:*
15: em_gstfun:*
16: em_gstcom:*
17: gb_gst8:*
18: em_gsthum:*
19: em_gstinv:*
20: em_gstpin:*
21: em_gstvt:*
22: em_gstfun:*
23: em_gstmam:*
24: em_gstmus:*
25: em_gstother:*
26: em_gstpro:*
27: em_gstrod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	100.0	501	AA211521	AA211521 zn55b01.r
2	149	100.0	683	BB556875	BB556875 7F70C03.x
3	149	100.0	719	BM697544	BM697544 UI-E-DXO-
4	149	100.0	756	BF791178	BF791178 602251278
5	149	100.0	793	BF671987	BF671987 602152408
6	149	100.0	826	BF126275	BF126275 601650456

7	149	100.0	879	12	BF575112	BF575112 602134792
8	149	100.0	906	12	BF672126	BF672126 602152580
9	147.4	98.9	909	12	BF790243	BF790243 602249777
10	144.8	97.2	855	12	BF672902	BF672902 602152759
11	138	92.6	745	17	A0751265	A0751265 HS 5574.B
12	137	91.9	330	9	A1355903	A1355903 BH11-F1-C
13	137	91.9	330	9	A1355904	A1355904 BH11-F1-C
14	133.8	89.8	683	12	BF111459	BF111459 7134h12.x
15	133	89.3	412	12	BG223652	BG223652 1M00010FO
16	133	89.3	517	12	BG224245	BG224245 1M0028R05
17	133	89.3	521	12	BF261282	BF261282 1M0004B05
18	133	89.3	528	12	BF073866	BF073866 220906.MA
19	133	89.3	582	12	BG224246	BG224246 1M0030E01
20	133	89.3	582	12	BG223758	BG223758 1M00018G0
21	127.2	85.4	466	12	BF358790	BF358790 BOVMS1-01
22	126.8	85.1	394	9	AA389647	AA389647 M104.Fera
23	123.6	83.0	499	12	BG224307	BG224307 1M0033A12
24	122.4	82.1	544	10	BE014221	BE014221 125832.MA
25	120.8	81.1	426	10	BE013988	BE013988 125544.MA
26	112.8	75.7	587	9	AA211443	AA211443 zn55b01.r
27	112	75.2	871	12	BF693124	BF693124 602080151
28	100.6	67.5	867	12	BF790636	BF790636 602250354
29	96.2	64.6	402	10	AW743181	AW743181 up63f10.y
30	96.2	64.6	465	14	W18646	W18646 mb98a05.r1
31	96.2	64.6	486	9	A1035961	A1035961 ub50D12.r
32	96.2	64.6	490	9	AA060214	AA060214 m137c11.r
33	96.2	64.6	504	9	AA033164	AA033164 H026C08-
34	96.2	64.6	613	14	BQ554133	BQ554133 UI-E-E30-
35	96.2	64.6	657	14	BQ176246	BQ176246 UTSM SMIG
36	96.2	64.6	890	9	AV088480	AV088480 AV088480
37	96.2	64.6	934	11	AK003105	AK003105 Mus muscu
38	96.2	64.6	934	11	AK010172	AK010172 Mus muscu
39	96.2	64.4	541	14	BM717052	BM717052 UI-E-E30-
40	96.2	64.4	544	14	BM674432	BM674432 UI-E-E30-
41	94.6	63.5	617	12	BG794218	BG794218 UTSM SMIG
42	94.4	63.4	482	9	AA434782	AA434782 ve23c01.r
43	90.8	60.9	505	9	AT878904	AT878904 H14E07.Y
44	87.4	58.7	578	14	B0554132	B0554132 H4026C08-
45	85.8	57.6	257	9	AA849490	AA849490 EST192257

ALIGNMENTS

RESULT 1
AA211521
LOCUS 501 bp mRNA linear EST 31-JAN-1997
DEFINITION zn55b01.r1 StrataGene muscle 937209 Homo sapiens CDNA clone
IMAGE:562057 5', mRNA sequence.
ACCESSION AA211521
VERSION AA211521.1 GI:1810175
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 501)
AUTHORS Hillier, L., Lemon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins,
'M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,
'B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevisan, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Warr, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the

COMMENT

TITLE JOURNAL MEDLINE
COMMENT

BASE COUNT	251 a	164 c	175 g	203 t
ORIGIN				
Query Match	100.0%;	Score 149;	DB 12;	Length 793;
Best Local Similarity	100.0%;	Pred. No. 4.7e-31;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1	GGGTGTTCCCTCCCACTCGATGAGAGAAAGCAATTCGAGAGCAAACTTCCA	60	
Db	279	GGGTGTTCTCTCCCACTCGATGAGAGAAAGCAATTCGAGAGCAAACTTCCA	338	
Oy	61	GGACCTGCAGTCAATCTATCGGAAATTCAGAAATTTAAAGTAACTAATATGTCCCC	120	
Db	339	GGACCTGCAGTCAATCTATCGGAAATTCGAAATTTAAAGTAACTAATATGTCCCC	398	
Oy	121	AAAGCTGAACAGTAGTAGGAGAAAGAAAAG	149	
Db	399	AAAGCTGAACAGTAGTAGGAGAAAGAAAAG	427	
RESULT 6				
BF126275	826 bp	mRNA	linear	EST 24-OCT-2000
DEFINITION	601850456p1 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:3934333 5',			
LOCUS	BF126275			
ACCESSION	BF126275			
VERSION	BF126275.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
TITLE	1 (bases 1 to 826)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: cga@db-remail.nih.gov			
	Tissue Procurement: CLONETECH Laboratories, Inc.			
	cDNA Library Preparation: CLONETECH Laboratories, Inc.			
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
	DNA Sequencing by: Incyte Genomics, Inc.			
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov			
	Plate: L1CMT74 row: D column: 14			
	High quality sequence stop: 651.			
FEATURES				
source				
	1..826			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:3934333"			
	/clone_1ib="NIH_MGC_76"			
	/lab_host="DH10B (TI phage-resistant)"			
	/note="Organ. liver; Vector: pDNR-LIB (Clontech); Site 1: SfiI (ggcgccctggcgc); Site 2: SfiI (ggcgcatatggc); 5' adaptor sequence: 5'-CACGGCCATTAGGCC-3' and 3' adaptor sequence 5'-ATTCTAGAGCCGACGCGCCGACCAATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.8 kb, (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC Library."			
BASE COUNT	280 a	156 c	188 g	202 t
ORIGIN				
Query Match	100.0%;	Score 149;	DB 12;	Length 826;
Best Local Similarity	100.0%;	Pred. No. 4.7e-31;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1	GGGTGTTCCCTCCCACTCGATGAGAGAAAGCAATTCGAGAGCAAACTTCCA	60	

Db	226	GGTGTCTCCTCCACCTCGGATAGAGAABAAGCAATTCACAGAGGAAGAAACTTCCA	285
OY	61	GGACCTGCAGTCAATCTATCGAATAATCTTAAGTGAACTAAATATGTCCCC	120
Db	286	GGACCTGCAGTCAATCTATCGAATAATCTTAAGTGAACTAAATATGTCCCC	345
OY	121	AAAGCTGAACGTAGTAGGAAGAAAAAG	149
Db	346	AAAGCTGAACGTAGTAGGAAGAAAAAG	374

RESULT 7
BF575112

LOCUS	BF575112	879 bp	mRNA	linear	EST 12-DEC-2000
DEFINITION	602134792F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4289735 5'				
ACCESSION	BF575112				
VERSION	BF575112.1 GI:11648824				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 879)				
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs@mail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Inyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: L1CM1131 row: h column: 24 High quality sequence stop: 595.				

FEATURES

source

Location/Qualifiers

1..879

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4289735"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccgtgcgc); Site_2: SfiI (ggccatcatgccc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCCGAGGCCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT	284 a	147 c	239 g	208 t	1 others
ORIGIN					

Query Match 100.0%; Score 149; DB 12; Length 879;
Best local Similarity 100.0%; Pred. No. 4.7e-31;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1	GGTGTCTCCTCCACCTCGGATAGAGAAGAACCAATTCACAGAGCGGAAGAACTTCCA	60
Db	277	GGTGTCTCCTCCACCTCGGATAGAGAAGAACCAATTCACAGAGCGGAAGAACTTCCA	336
OY	61	GGACCTGCAGTCAATCTATCGAATAATCTTAAGTGAACTAAATATGTCCCC	120
Db	337	GGACCTGCAGTCAATCTATCGAATAATCTTAAGTGAACTAAATATGTCCCC	396
OY	121	AAAGCTGAACGTAGTAGGAAGAAAAAG	149

DB 397 AAAGCTGAACAGTAGTAGGAAGAAAAG 425

RESULT 8
LOCUS BF672126
DEFINITION 602152580P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293763 5',
MRNA sequence.
ACCESSION BF672126 906 bp mRNA linear EST 21-DEC-2000
VERSION BF672126.1 GI:11946021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 906)
NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LNCM1141 row: P column: 20
High quality sequence stop: 621.
Location/Qualifiers
1..906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4293763"
/clone_11b="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site 1: SfiI (ggcgccctcgsc); Site 2: SfiI
(ggccatcagcgc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

BASE COUNT 285 a 208 c 200 g 213 t

ORIGIN

Query Match 100.0%; Score 149; DB 12; Length 906;
Best Local Similarity 100.0%; Pred. No. 4.7e-31;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGTTCTCTCCACCTCGATGAGAGAAAGCAATTCAGAGCGAAGAACTTCCA 60
DB 246 GGTGTTCTCTCCACCTCGATGAGAGAAAGCAATTCAGAGCGAAGAACTTCCA 305

QY 61 GGAACCTGCAGTCAATCTATCGAAATCCAGAAATTTAAAGTAATTAATATGTCCCC 120
DB 306 GGAACCTGCAGTCAATCTATCGAAATCCAGAAATTTAAAGTAATTAATATGTCCCC 365

QY 121 AAAGCTGAACAGTAGTAGGAAGAAAAG 149
DB 366 AAAGCTGAACAGTAGTAGGAAGAAAAG 394

RESULT 9
LOCUS BF790243 909 bp mRNA linear EST 12-JAN-2001
DEFINITION 602249777P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4328154 5',
MRNA sequence.

ACCESSION BF790243
VERSION BF790243.1 GI:12095188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 909)
NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: LNCM1187 row: I column: 19
High quality sequence stop: 626.
Location/Qualifiers
1..909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328154"
/clone_11b="NIH MGC 81"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site 1: SfiI (ggcgccctcgsc); Site 2: SfiI
(ggccatcagcgc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."

BASE COUNT 279 a 200 c 212 g 218 t

ORIGIN

Query Match 98.9%; Score 147.4; DB 12; Length 909;
Best Local Similarity 99.3%; Pred. No. 1.3e-30;
Matches 148; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGTGTTCTCTCCACCTCGATGAGAGAAAGCAATTCAGAGCGAAGAACTTCCA 60
DB 315 GGTGTTCTCTCCACCTCGATGAGAGAAAGCAATTCAGAGCGAAGAACTTCCA 374

QY 61 GGAACCTGCAGTCAATCTATCGAAATCCAGAAATTTAAAGTAATTAATATGTCCCC 120
DB 375 GGAACCTGCAGTCAATCTATCGAAATCCAGAAATTTAAAGTAATTAATATGTCCCC 434

QY 121 AAAGCTGAACAGTAGTAGGAAGAAAAG 149
DB 435 AAAGCTGAACAGTAGTAGGAAGAAAAG 463

RESULT 10
LOCUS BF672902 855 bp mRNA linear EST 21-DEC-2000
DEFINITION 602152759P1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4293878 5',
MRNA sequence.
ACCESSION BF672902
VERSION BF672902.1 GI:11946797
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 855)

ccq. . T

BASE COUNT	270 a	159 c	210 g	216 t
ORIGIN				

scanning the human genome
JOURNAL, Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L

FEATURES

Source

BASE COUNT	245 a	132 c	139 g	228 t	1 others
ORIGIN					

DQ 120 CAAAGCTGAACAGTAGTAGCAAGAAAAAAG 148
 |||||
Db 109 CAAAGCTGAACAGTAGTAGCAAGAAAAAAG 80

Department of Molecular Human Genetics
Institut für Human Genetics
Fleischmannstr. 42/44, D-17487 Greifswald, Germany
Tel.: +49 3854 8653-78 (-74)

Email: patzak@rz.uni-greifswald.de (mwehnert@rz.uni-greifswald.de)


```

source
1. .683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3523462"
/lab_host="Soares NSF_F8_9W_OT_PA_P_S1"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were made in vitro.
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clonesID: Soares NbHSP pool 1:
309384-310919, 323208-325895 Soares NbZHP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NbHSP-9W pool 1:
758280-760583, 772104-774407 Soares NbHSP pool 1:
304776-306311, 320136-322823, 326280-326653 Soares NbHOT
pool 1: 723720-726407, 739080-740999 Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT      199 a      127 c      140 g      217 t
ORIGIN
Query Match      89.8%; Score 133.8; DB 12; Length 683;
Best Local Similarity 95.2%; Pred. No. 7.4e-27;
Matches 138; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 TTCTCCACCTCGGATGAGAGAGAGAGCAATTCGAGGCGAGAGAACTTCAGAGC 64
DB 606 TCTCTCCACATCGCATGATGAGAGAGAGAGCAATTCGAGGCGAGAGAACTTCAGAGC 547
QY 65 CTGCGATCAATCTATCGGAATTCGAGATATTAAAGTGAATCTAAATATGTCCTCCCAAG 124
DB 546 CTGCGATCAATCTATCGGAATTCGAGATATTAAAGTGAATCTAAATATGTCCTCCCAAG 487
QY 125 CTGACAGTAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 149
DB 486 CTGACAGTAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462

RESULT 15
BG223652      412 bp      mRNA      linear      EST 07-FEB-2001
LOCUS      1M00010F05 Bovine Mixed Skeletal Muscle cDNA Library Bos taurus
DEFINITION      cDNA 5', mRNA sequence.
ACCESSION      BG223652
VERSION      BG223652.1 GI:12709185
KEYWORDS      EST.
SOURCE      cow.
ORGANISM      Bos taurus
            Eukaryota; Metazoa; Chordata; Graniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
            Bovidae; Bovinae; Bos.
            1 (bases 1 to 412)
            Moore,S.S., Hansen, C., Li,C., Fu,A., Meng,Y., Li,G., Murdoch,G.,
            Dixon,W. and Christopherson,B.
            cDNA's from bovine mixed skeletal muscle
            Unpublished (2001)
            Contact: Dr. Stephen S. Moore
            Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410-Agril/For Centre, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5,
            Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoores@afns.ualberta.ca
            The sequence best matches gb:U84250584 (Homo sapiens mRNA for
            stretch responsive muscle (X-chromosome) protein (strmx gene).
            9/2000 ) in GenBank main database at E-value of 1e-115.
            PCR Primers
            FORWARD: M13 Forward

```

```

FEATURES
SOURCE
location/Qualifiers
1. .412
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Mixed Skeletal Muscle cDNA Library"
/sex="Two males and one female mixed"
/tissue_type="masseter, longissimuslorsi, biceps femoris,
semimembranes"
/cell_type="muscle"
/dev_stage="young adult"
/lab_host="X11-BiueKRF-strain"
/notes="Organ: Skeletal muscle; Vector: Uni-22APXR; Site_1:
Ecor I; Site_2: Xho I"

BASE COUNT      147 a      83 c      97 g      85 t
ORIGIN
Query Match      89.3%; Score 133; DB 12; Length 412;
Best Local Similarity 93.3%; Pred. No. 1.2e-26;
Matches 139; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GGTGTTCTCCACCTCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
DB 143 GGTGTTCTCCACCTCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 202
QY 61 GGACCTGACATCAATCTATCGGAATTCGAGATATTAAAGTGAATCTAAATATGTCCTCC 120
DB 203 GGACCTGACATCAATCTATCGGATTCAGAAACATTAAGTGAATCTAAATATGTCCTCC 262
QY 121 AAAGCTGAACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 149
DB 263 AAAGCTGAACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291

Search completed: April 15, 2003, 22:02:35
Job time : 312.035 secs

```


GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 15, 2003, 18:42:01 ; Search time 8.81622 Seconds
(without alignments)
5183.040 Million cell updates/sec

Title: US-09-647-019-9

Perfect score: 149
Sequence: 1 gggtgtctctccacccgcga.....cagtagtagaagaagaaag 149

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCCTUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/beckfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	100.0	909	4 US-09-484-970B-111	Sequence 111, App
2	29.6	19.9	2973	4 US-09-133-962A-15	Sequence 15, Appl
3	29.2	19.6	2133	4 US-09-164-034B-1	GENERAL INFORMA
4	29.2	19.6	2653	1 US-08-325-553-1	Sequence 1, Appl
5	29.2	19.6	2653	2 US-08-394-152A-1	Sequence 1, Appl
6	28.8	19.3	2919	4 US-09-710-092-1	Sequence 5, Appl
7	28.8	19.3	3171	4 US-09-710-092-5	Sequence 9, Appl
8	28.8	19.3	3813	4 US-09-710-092-9	Sequence 13, Appl
9	28.8	19.3	4281	4 US-09-710-092-13	Sequence 17, Appl
10	28.8	19.3	5358	4 US-09-710-092-17	Sequence 1, Appl
11	28.4	19.1	51952	3 US-08-947-823-1	Sequence 6, Appl
12	28	18.8	2073	4 US-09-032-523-6	Sequence 18, Appl
13	27.6	18.5	33000	4 US-09-215-694-18	Sequence 10, Appl
14	27.6	18.5	98844	4 US-09-791-211-10	Sequence 8, Appl
15	27.2	18.3	1267	3 US-08-872-979-2	Sequence 211, App
16	27.2	18.3	1267	3 US-08-872-979-2	Sequence 2, Appl
17	27	18.1	1796	1 US-08-548-509-1	Sequence 1, Appl
18	27	18.1	3810	2 US-08-475-844-8	Sequence 8, Appl
19	27	18.1	4156	5 US-08-961-527-211	Sequence 8, Appl
20	27	18.1	5835	4 US-09-033-333-3	Sequence 211, App
21	27	18.1	5835	4 US-09-033-333-3	Sequence 2, Appl
22	27	18.1	5835	4 US-09-614-495-3	Sequence 3, Appl
23	27	18.1	5836	1 US-08-380-916-1	Sequence 2, Appl
24	27	18.1	5836	3 US-08-721-690-1	Sequence 3, Appl
25	27	18.1	5836	4 US-08-891-581-1	Sequence 1, Appl
26	27	18.1	5836	4 US-09-033-333-2	Sequence 2, Appl
27	27	18.1	5836	4 US-09-033-333-2	Sequence 1, Appl

28	27	18.1	5836	4 US-09-614-495-2	Sequence 2, Appl
29	27	18.1	5827	4 US-09-813-133A-3	Sequence 3, Appl
30	26.8	18.0	605	1 US-08-592-126-77	Sequence 77, Appl
31	26.8	18.0	1929	4 US-09-178-252-24	Sequence 24, Appl
32	26.8	18.0	1975	4 US-09-328-571A-12	Sequence 12, Appl
33	26.8	18.0	2118	4 US-09-134-001C-2519	Sequence 2519, App
34	26.8	18.0	2539	3 US-08-749-522-3	Sequence 3, Appl
35	26.8	18.0	3558	4 US-09-178-252-22	Sequence 3, Appl
36	26.6	17.9	1113	4 US-09-134-001C-1329	Sequence 1329, App
37	26.6	17.9	5115	3 US-08-348-518C-3	Sequence 3, Appl
38	26.6	17.9	5115	3 US-08-476-509B-3	Sequence 3, Appl
39	26.6	17.9	168575	4 US-09-426-290-1	Sequence 1, Appl
40	26.4	17.7	405	4 US-09-134-001C-23	Sequence 1, Appl
41	26.4	17.7	735	2 US-08-743-637B-194	Sequence 194, Appl
42	26.4	17.7	828	4 US-09-693-147-1	Sequence 12, Appl
43	26.4	17.7	857	4 US-08-460-040-1	Sequence 1, Appl
44	26.4	17.7	860	4 US-08-858-207A-172	Sequence 172, App
45	26.4	17.7	1172	1 US-07-945-288-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-484-970B-111
; Sequence 111, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 111
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 021656.2CB1
US-09-484-970B-111

Query Match 100.0%; Score 149; DB 4; Length 909;
Best Local Similarity 100.0%; Pred. No. 1e-38;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTTGTTCTCCCACTCGGATGAGAGAAAGCCAAATTCGAGAGGAGAACTTCCA 60
DB 354 GGTTGTTCTCCCACTCGGATGAGAGAAAGCCAAATTCGAGAGGAGAACTTCCA 413
QY 61 GGACCTGACGATCATCTGGAATCCAGATATTAAGGAACTAAATATGTCCCC 120
DB 414 GGACCTGACGATCATCTGGAATCCAGATATTAAGGAACTAAATATGTCCCC 473
QY 121 AAAGCTGAACGATGAGAGAAAGGAAAAAG 149
DB 474 AAAGCTGAACGATGAGAGAGAAAAAG 502

RESULT 2
US-09-133-962A-15/c
; Sequence 15, Application US/09133962A
; Patent No. 6372965
; GENERAL INFORMATION:
; APPLICANT: JONATHAN EDWARD LIGHTNER
; TITLE OF INVENTION: GENES FOR MICROSOAL FATTY ACID
DELTA-12 DESATURASES AND RELATED
ENZYMES FROM PLANTS

NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/133,962A
FILING DATE: 14-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 07/977,339
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHRISTENBURY, LYNN M.
REGISTRATION NUMBER: 30,971
REFERENCE/DOCKET NUMBER: BB-1043-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-5481
TELEFAX: (302) 773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arabidopsis thaliana
IMMEDIATE SOURCE:
CLONE: pAGF2-6
FEATURE:
NAME/KEY: exon
LOCATION: 433..520
FEATURE:
NAME/KEY: intron
LOCATION: 521..1654
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-133-962A-15
Query Match 19.9%; Score 29.6; DB 4; Length 2973;
Best Local Similarity 54.6%; Pred. No. 2.3;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 26 AGAAGAGCCGATTCGAGAGCGAAGAACTTCGAGAGCTGAGTCATCTATCGGAA 85
DB 1259 AAAAGAAAAGCTGAGTGTATGATTAAGCCAGCAAGATTGTGTGACATCTATATCA 1200
QY 86 TCCGAAATATTAAGTGAAGTAATAATATGTCCCAAGCTGAACAGT 133
DB 1199 TTTAGATATTTAAACAGAAATTAATAATGTGACGAGGATCTAGAAAAT 1152
RESULT 3
US-09-164-034B-1
GENERAL INFORMATION:
APPLICANT: Mincheff, Milcho S.
Lounkinov, I. Dmitri
Zoubak, Serguei
TITLE OF INVENTION: Immunotherapy of Cancer Through Expression
of Truncated Tumor- or Tumor-Associated Antigen
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:

ADDRESSEE: William S. Ramsey,
Ramey, Cook, Looper & Kurlander, LLC
STREET: 10420 Little Patuxent Parkway, Suite 250
CITY: Columbia
STATE: Maryland
COUNTRY: USA
ZIP: 21044
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
COMPUTER: PC
OPERATING SYSTEM: Windows 95
SOFTWARE: Wordperfect 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/164,034B
FILING DATE: 30-Sep-1998
ATTORNEY/AGENT INFORMATION:
NAME: Ramsey, William S.
REGISTRATION NUMBER: 32,715
REFERENCE/DOCKET NUMBER: br11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (410) 992-9660
TELEFAX: (410) 992-9540
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-164-034B-1
Query Match 19.6%; Score 29.2; DB 4; Length 2133;
Best Local Similarity 50.7%; Pred. No. 2.7; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 6 TCCCTCCACCCTCGATGAGAGAGAGGCAATTCGAGAGCGAAGAACTTCGAGACC 65
DB 567 TGCTGACTACTTGTCTCCGGGGTGAAGTCCATCAAGATGTTGGAATCTTCGAGG 626
QY 66 TCGATCATCTATCGGAATTCAGAAATTTAAAGTGAACCTAAATATGTCCCAAGC 125
DB 627 TGGTGTCCAGCGGAGAAATCTTAATCTGAATGTGAGGAGACCTTCACACAGG 686
QY 126 TGAACAGTAGTAGGAGA 143
DB 687 TTACCCAGCAATGAAATTA 704
RESULT 4
US-08-325-553-1
Sequence 1, Application US/0832553
Patent No. 5538866
GENERAL INFORMATION:
APPLICANT: Israel, Ron S.
APPLICANT: Heston, Warren D.W.
APPLICANT: Fair, William R.
TITLE OF INVENTION: THE PROSTATE-SPECIFIC MEMBRANE ANTIGEN
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/973,337A
FILING DATE: 05 NOV 1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

```

1      REGISTRATION NUMBER: 28, 678
2      REFERENCE/DOCKET NUMBER: 1747/41426
3      TELECOMMUNICATION INFORMATION:
4      TELEPHONE: (212) 977-9550
5      TELEFAX: (212) 664-0525
6      TELEX: 422523 COOP UI
7      INFORMATION FOR SEQ ID NO: 1:
8      SEQUENCE CHARACTERISTICS:
9      LENGTH: 2653 base pairs
10     TYPE: nucleic acid
11     STRANDEDNESS: double
12     TOPOLOGY: linear
13     MOLECULE TYPE: cDNA
14     HYPOTHETICAL: NO
15     ANTI-SENSE: NO
16     ORIGINAL SOURCE:
17     ORGANISM: Homo sapiens
18     TISSUE TYPE: Carcinoma
19     IMMEDIATE SOURCE:
20     CLONE: Prostate-Specific Membrane Antigen
21     FEATURE:
22     NAME/KEY: CDS
23     LOCATION: 262..2511
24     US-08-325-553-1

```

	Query Match	Best Local Similarity	Score 29.2;	DB 1;	Length 2653;
	Matches 70;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0
Qy	6	TCCCTCCACCTTGGATGAGAGAGAGAGCCCAATTCGAGAGCGAGAGAACTTCCAGACC	65		
Db	954	TGCTGACTACTTTGCTCCCGGGGTGAAGTCCTATCCAGATGGTTGGAATCTTCTCTGGAG	1013		
Qy	66	TGCACTCAATCTTATGGAGAAATCCAGAAATTAAAGTGAACCTAAATATATGTCCCAAGC	125		
Db	1014	TGGTCTCCACGCTGGAGAAATATCTCTTAATCTGAAATGCTGACGAGAGACCTCTCACACGAG	1073		
Qy	126	TGAACAGTAAGTAGAGAGA	143		
Db	1074	TTACCAGCAATGAGATA	1091		

RESULT 5
 US-08-394-152A-1
 : Sequence 1, Application US/08394152A
 : Patent No. 5935818
 :
 : GENERAL INFORMATION:
 : APPLICANT: Israel1, Ron S.
 : APPLICANT: Heaton, Warren D.W.
 : APPLICANT: Fair, William R.
 : TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
 : TITLE OF INVENTION: USES THEREOF
 : NUMBER OF SEQUENCES: 48
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 :
 : STATE: New York
 :
 : COUNTRY: United States of America
 : ZIP: 10036
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM 330 466 DX2
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/394,152A
 : FILING DATE: 24-FEB-95
 : CLASSIFICATION: 435
 :
 : ATTORNEY/AGENT INFORMATION:
 : NAME: White, John P.
 : REGISTRATION NUMBER: 28,678
 :
 : REFERENCE/DOCKET NUMBER: 41426-B

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2653 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Carcinoma
IMMEDIATE SOURCE:
CLONE: Prostate-Specific Membrane Antigen
FEATURE:
NAME/KEY: CDS
LOCATION: 262..2511

```

Query Match	19.6%	Score 29.2	DB 2	Length 2653
Best Local Similarity	50.7%	Pred. No. 2.9		
Matches	70	Conservative	0	Mismatches 68; Indels 0; Gaps 0
QY	6	TCCTCCACCTCGATGAGAGAAAGCCCAATTCGAGAGCGAAGAACTTCCAGACC	65	
Db	954	TGCTACACTTCTTGCTCCTGGGGTGAAAGTCCATCCAGATGGTTGGAACTTCCCTGGAGG	1013	
QY	66	TGCAGTCAATCTANTCGAAATCCAGAAATATTTAAAGTGAAGTAAATATGTCCCAAGC	125	
Db	1014	TGGTGTCCAGCGGTGGAATATCTCTAAATCTGATGATGTGACAGAGACCTCTCACACGAG	1073	
QY	126	TGAACAGTAGTAGAAGA	143	
Db	1074	TTACCCAGCAAAATGATA	1091	

```

RESULT 6
US-09-710-092-1
; Sequence 1, Application US/09710092
; Patent No. 6462186
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Hilbun, Erin
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6462186el Human Arpase Proteins and
; TITLE OF INVENTION: Polynucleotide Encoding the Same
; FILE REFERENCE: LEX-0085-USA
; CURRENT APPLICATION NUMBER: US/09/710.092
; CURRENT FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/164,624
; PRIOR FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2919
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-710-092-1

Query Match          19.3%; Score 28.8; DB 4; Length 2919;
Best Local Similarity 62.5%; Pred. No. 4;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0

QY      77  TATCGGAATTCGAGATATTAAAGTGAAGCTAAATATGTCCCCCAAGCTGAACGTAAGT 136
Db      427 TATCGGAATTCGAGATATTAAAGTGAAGCTAAATATTAAATTAAGCTTAAGTATAGTAGG 486

```

QY 137 AGGAGAGAAAA 148
Db 487 AAAGAGAAAAA 498

RESULT 7

US-09-710-092-5
; Sequence 5, Application US/09710092
; Patent No. 6462186
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Hilbun, Erin
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Sambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6462186e1 Human ATPase Proteins and
; FILE REFERENCE: LEX-0085-USA
; CURRENT APPLICATION NUMBER: US/09/710,092
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/164,624
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-710-092-5

Query Match 19.3%; Score 28.8; DB 4; Length 3171;
Best Local Similarity 62.5%; Pred. No. 4.2;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 77 TATCGGAATCCAGATATTAAGTAACTAATATGTCCTCAAGCTGAACAGTAGT 136
Db 427 TATCGGAATACAAATTTGCAACAGATCAATTAATTAAGTTTATAGTAGG 486

QY 137 AGGAGAGAAAA 148
Db 487 AAAGAGAAAAA 498

RESULT 8

US-09-710-092-9
; Sequence 9, Application US/09710092
; Patent No. 6462186
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Hilbun, Erin
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Sambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6462186e1 Human ATPase Proteins and
; FILE REFERENCE: LEX-0085-USA
; CURRENT APPLICATION NUMBER: US/09/710,092
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/164,624
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 3813
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-710-092-9

Query Match 19.3%; Score 28.8; DB 4; Length 3813;
Best Local Similarity 62.5%; Pred. No. 4.4;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 77 TATCGGAATCCAGATATTAAGTAACTAATATGTCCTCAAGCTGAACAGTAGT 136
Db 427 TATCGGAATACAAATTTGCAACAGATCAATTAATTAAGTTTATAGTAGG 486

QY 137 AGGAGAGAAAA 148
Db 487 AAAGAGAAAAA 498

RESULT 9

US-09-710-092-13
; Sequence 13, Application US/09710092
; Patent No. 6462186
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Hilbun, Erin
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Sambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6462186e1 Human ATPase Proteins and
; FILE REFERENCE: LEX-0085-USA
; CURRENT APPLICATION NUMBER: US/09/710,092
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/164,624
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 4281
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-710-092-13

Query Match 19.3%; Score 28.8; DB 4; Length 4281;
Best Local Similarity 62.5%; Pred. No. 4.6;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 77 TATCGGAATCCAGATATTAAGTAACTAATATGTCCTCAAGCTGAACAGTAGT 136
Db 427 TATCGGAATACAAATTTGCAACAGATCAATTAATTAAGTTTATAGTAGG 486

QY 137 AGGAGAGAAAA 148
Db 487 AAAGAGAAAAA 498

RESULT 10

US-09-710-092-17
; Sequence 17, Application US/09710092
; Patent No. 6462186
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Hilbun, Erin
; APPLICANT: Nehls, Michael C.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Sambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6462186e1 Human ATPase Proteins and
; FILE REFERENCE: LEX-0085-USA
; CURRENT APPLICATION NUMBER: US/09/710,092
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: US 60/164,624
; NUMBER OF SEQ ID NOS: 17

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 17
LENGTH: 5958
TYPE: DNA
ORGANISM: homo sapiens
US-09-710-092-17

Query Match 19.3% Score 28.8; DB 4; Length 5958;
Best Local Similarity 62.5%; Pred. No. 5.1;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 77 TATGGAAATTCAGAAATTTAAAGTCACTAAATATGTCCTCCCAAGCTGAACGTAAGT 136
DB 464 TATGGAAATTCAGAAATTTGACAAACAGATCAATATTAATTAAGTTTATATGAG 523
QY 137 AGAAGAAAAA 148
DB 524 AAGAGAAAAA 535

RESULT 11
US-08-947-823-1

Sequence 1, Application US/08947823

Patent No. 6114605

GENERAL INFORMATION:

APPLICANT: Williamson, Valerie M.

APPLICANT: Kaloshian, Isagouni

APPLICANT: Vaghoobi, Jafar

APPLICANT: Bodeau, John

APPLICANT: Milligan, Stephen

TITLE OF INVENTION: Procedures and Materials for Confering

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/947,823

FILING DATE: 09-OCT-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/18802

FILING DATE: 09-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/028,191

FILING DATE: 10-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Baetian, Kevin L.

REGISTRATION NUMBER: 34,774

REFERENCE/DOCKET NUMBER: 023070-070210US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 51952 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-947-823-1

Query Match 19.1% Score 28.4; DB 3; Length 51952;

Best Local Similarity 53.6%; Pred. No. 14;

Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 35 CAATTCAGAGGAGGAGAAATCTCCAGAGCTGCAATCTATCGAAATCCAGATA 94

DB 6001 CAGTTCCTAATGTGCAAACTAGCAACAACTCATGCTATTCAGCAATCAAAATA 6060

QY 95 TTAAGTGAAGTAAATATGTCCTCCCAAGCTGAACGTAGTAGAAGAA 144

DB 6061 CAATACACACATTTATATGACCAAAATTTGAAAGCTGAACCCAAA 6110

RESULT 12
US-09-032-523-6

Sequence 6, Application US/09032523

Patent No. 6233454

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Hillman, Jennifer L.

APPLICANT: Corley, Neil C.

APPLICANT: Guegler, Karl

APPLICANT: Baugh, Mariah

TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/032,523

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0479 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2073 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: PANTCT01

CLONE: 1515165

US-09-032-523-6

Query Match 18.8% Score 28; DB 4; Length 2073;

Best Local Similarity 58.3%; Pred. No. 6.5;

Matches 49; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 65 CTGCACTCAATCTATCGAATTCAGAAATATTAAGTGAAGTAAATATGTCCTCCCAAG 124

DB 1975 CTTATATCATTTATTCAGAAAGTTCAAGCATTTTAAAAAAGAAAAAAGAAAAAACA 2034

QY 125 CTGAACAGTAGTAGAAGAAAAA 148

DB 2035 AAGAAAAAAGAAAAAAGAAAAA 2058

```

RESULT 13
US-09-215-694-18
; Sequence 18, Application US/09215694B
; Patent No. 6391583
; GENERAL INFORMATION:
; APPLICANT: Wisconsin Alumni Research Foundation
; APPLICANT: Hutchinson, Charles R.
; APPLICANT: Kennedy, Jonathan n.m.i
; APPLICANT: Park, Cheonsuek n.m.i
; TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS
; FILE REFERENCE: 960296.95718
; CURRENT APPLICATION NUMBER: US/09/215,694B
; CURRENT FILING DATE: 1999-12-18
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 18
; LENGTH: 33000
; TYPE: DNA
; ORGANISM: Aspergillus terreus
US-09-215-694-18

Query Match      18.5%; Score 27.6; DB 4; Length 33000;
Best Local Similarity 63.6%; Fred.No.22;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0

Oy      143 AAAAAA 148
Db      30166 AAAATATAAAATTATAAGTGAACCTAAATATGTCCCCAAAGCTGAACGATAGTAGAG 142
        ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
        ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

RESULT 14
US-09-791-211-10
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Walt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-0205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 98844
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 24962
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 64383
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65468
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65469
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65470
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65471
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 87130
; OTHER INFORMATION: unknown

```

```

/ NAME/KEY: unsure
/ LOCATION: 89049
/ OTHER INFORMATION: unknown
/ OTHER INFORMATION:
US-09-791-211-10

Query Match      18.5%; Score 27.6; DB 4; Length 98844;
Best Local Similarity 53.8%; Pred. No. 32;
Matches 57; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY      8 CTCGCACCTCGAATGAGGAGAAAGCAATTCACGAGCGAAGAACTTCACGAGCCTG 67
Db      93147 CTGATACCAGGCTGATGATGCGGCCAATCCGGCTCACTACATCTTCACCTCTG 93206
        |||||
OY      68 CAGTCATCTATCGGAATCCGAAATCCAGATATTTAAAGTGAAGTAAGTAATA 113
Db      93207 GGTTCAGACTATTCCTCGCTCAGTCTCCAGAGTGAATTA 93252
        |||||

RESULT 15
US-08-872-979-2/c
; Sequence 2, Application US/08872979
; Patent No. 6074844
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: TWO NEW HUMAN MEMBRANE FUSION PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/872,979
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0320 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1267 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTN0703
; CLONE: 1363873
US-08-872-979-2

Query Match      18.3%; Score 27.2; DB 3; Length 1267;
Best Local Similarity 51.7%; Pred. No. 9.9;
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY      14 CTCGCATGAGGAGGAAGCAATTCACGAGCGAAGAACTTCACAGACTCGAGTCA 73
        |||||

```

us-09-647-019-9.rn1

Page 7

Search completed: April 15, 2003, 22:53:28
Job time : 62.8162 secs

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: April 15, 2003, 19:13:41 ; Search time 14.8883 Seconds

(without alignments)
8778.558 Million cell updates/sec

Title: US-09-647-019-9

Perfect score: 149
Sequence: 1 ggtcttccccaccctcgga.....cagtagagagaagaaag 149

Scoring table:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 593429 seqs, 438593890 residues

Total number of hits satisfying chosen parameters: 1186658

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodaca/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodaca/2/pubpna/PTC_NEW_PUB.seq:*
3: /cgn2_6/ptodaca/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodaca/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodaca/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodaca/2/pubpna/PTCUS_PUBCOMB.seq:*
7: /cgn2_6/ptodaca/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodaca/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodaca/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodaca/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodaca/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodaca/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodaca/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodaca/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	100.0	824	10	US-09-880-192-4
2	127.2	85.4	466	10	US-09-960-352-5216
3	112.8	75.7	587	10	US-09-962-436-468
4	36.6	24.6	1875	9	US-09-938-842A-3453
5	32.8	22.0	640681	10	US-09-790-988-1
6	31.8	21.3	2780	10	US-09-841-786-10
7	31.8	21.3	9726	10	US-09-841-786-10
8	31.8	21.3	11330	10	US-09-841-786-15
9	30	20.0	873	10	US-09-815-242-6627
10	29.8	20.0	32190	10	US-09-841-786-15
11	29.6	19.9	425	10	US-09-764-887-308
12	29.4	19.7	329	9	US-09-803-719-278
13	29.4	19.7	2232	9	US-10-087-464-45
14	29.2	19.6	477	10	US-09-864-761-14091
15	29.2	19.6	2558	9	US-09-878-295A-617
16	29.2	19.6	2558	9	US-09-878-697-617
17	29.2	19.6	2558	9	US-09-978-192A-617
18	29.2	19.6	2558	9	US-09-999-832A-617
19	29.2	19.6	2558	9	US-09-978-189-617

20	29.2	19.6	2558	9	US-10-174-590-103	Sequence 103, App
21	29.2	19.6	2558	9	US-10-176-758-103	Sequence 103, App
22	29.2	19.6	2558	9	US-10-175-737-103	Sequence 103, App
23	29.2	19.6	2558	9	US-10-173-706-103	Sequence 103, App
24	29.2	19.6	2558	9	US-10-175-738-103	Sequence 103, App
25	29.2	19.6	2558	9	US-10-175-732-103	Sequence 103, App
26	29.2	19.6	2558	9	US-10-176-482-103	Sequence 103, App
27	29.2	19.6	2558	9	US-10-176-757-103	Sequence 103, App
28	29.2	19.6	2558	9	US-10-176-757-103	Sequence 103, App
29	29.2	19.6	2558	9	US-10-180-552-103	Sequence 103, App
30	29.2	19.6	2558	9	US-10-180-557-103	Sequence 103, App
31	29.2	19.6	2558	9	US-10-173-700-103	Sequence 103, App
32	29.2	19.6	2558	9	US-10-174-572-103	Sequence 103, App
33	29.2	19.6	2558	9	US-10-174-579-103	Sequence 103, App
34	29.2	19.6	2558	9	US-10-174-582-103	Sequence 103, App
35	29.2	19.6	2558	9	US-10-174-588-103	Sequence 103, App
36	29.2	19.6	2558	9	US-10-175-739-103	Sequence 103, App
37	29.2	19.6	2558	9	US-10-175-740-103	Sequence 103, App
38	29.2	19.6	2558	9	US-10-175-743-103	Sequence 103, App
39	29.2	19.6	2558	9	US-10-176-488-103	Sequence 103, App
40	29.2	19.6	2558	9	US-10-176-492-103	Sequence 103, App
41	29.2	19.6	2558	9	US-10-176-747-103	Sequence 103, App
42	29.2	19.6	2558	9	US-10-176-750-103	Sequence 103, App
43	29.2	19.6	2558	9	US-10-176-985-103	Sequence 103, App
44	29.2	19.6	2558	9	US-10-176-987-103	Sequence 103, App
45	29.2	19.6	2558	9	US-10-176-991-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-880-192-4
Sequence 4, Application US/09880192
Patent No. US20020077470A1
GENERAL INFORMATION:
APPLICANT: Walker, Michael G.
APPLICANT: Volkman, Wayne
APPLICANT: Klingler, Tod M.
APPLICANT: Azimza, Yalda
TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
FILE REFERENCE: PB-0009-1 CIP
CURRENT APPLICATION NUMBER: US/09/880,192
CURRENT FILING DATE: 2001-06-12
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PERL Program
SEQ ID NO 4
LENGTH: 824
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20020077470A1 3601719CBI
US-09-880-192-4

Query Match 100.0%; Score 149; DB 10; Length 824;
Best Local Similarity 100.0%; Fied. No. 4,3e-37;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGTTCTCCACCTCGATGAGAGAGCAATTCAGAGGAGGAATTCCTCA 60
DB 350 GGTGTTCTCCACCTCGATGAGAGAGCAATTCAGAGGAGGAATTCCTCA 409
QY 61 GACCTGAGAGCAATTCATTCGAAATTCAGAAATTTAAAGTAACTAAATATGTCCTCC 120
DB 410 GACCTGAGAGCAATTCATTCGAAATTCAGAAATTTAAAGTAACTAAATATGTCCTCC 469
QY 121 AAAGCTGAACAGTAGTGAAGAAAGAAAG 149
DB 470 AAAGCTGAACAGTAGTGAAGAAAGAAAG 498

RESULT 2

US-09-790-988-1

Query Match 22.0%; Score 32.8; DB 10; Length 640681;

Best Local Similarity 58.0%; Pred. No. 12;

Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 49 AGAAGCTTCAGAGACCTGAGTCAATCTATCGGAAATCCAGAAATATTAAGTAAAG 108

DB 520435 AATTAAGATACAGAACTGATGTTGTTTATACAGCAACGCTATTAATATGACTA 520494

QY 109 AATATGTCCCAAGCTGAACAGTACTAGAGAGAAAA 148

DB 520495 ATAGATTATTTATTTATCTCAGCGTCAATAAAAATAAA 520534

RESULT 6

US-09-841-786-10

Sequence 10, Application US/09841786

Patent No. US20020054883A1

GENERAL INFORMATION:

APPLICANT: NAGARAJA, T. G.

APPLICANT: STEWART, GEORGE C.

APPLICANT: NARAYANAN, SANJEEV K.

TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN

FILE REFERENCE: 30296

CURRENT APPLICATION NUMBER: US/09/841,786

PRIOR FILING DATE: 2001-04-24

PRIOR APPLICATION NUMBER: 09/558,257

NUMBER OF SEQ ID NOS: 15

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 10

LENGTH: 2780

TYPE: DNA

ORGANISM: Fusobacterium necrophorum

US-09-841-786-10

Query Match 21.3%; Score 31.8; DB 10; Length 2780;

Best Local Similarity 56.1%; Pred. No. 3;

Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 42 AGAGCGAAGAACTTCAGACCTGAGTCAATCTATCGGAAATCCAGAAATTAAG 101

DB 1164 AGAAGAAAGAACTTATGGAATGAGAGAGTTTATGGAAGCTATTAATATATGC 1223

QY 102 TGAAGTAAATATGTCCCAAGCTGAACAGTACTAGAGAGAAAA 148

DB 1224 TTCTGTGACAAATGCCGATCATGCTGAATGGTATCGAAGAAAA 1270

RESULT 7

US-09-841-786-8

Sequence 8, Application US/09841786

Patent No. US20020054883A1

GENERAL INFORMATION:

APPLICANT: NAGARAJA, T. G.

APPLICANT: STEWART, GEORGE C.

APPLICANT: NARAYANAN, SANJEEV K.

TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN

FILE REFERENCE: 30296

CURRENT APPLICATION NUMBER: US/09/841,786

PRIOR FILING DATE: 2001-04-24

PRIOR APPLICATION NUMBER: 09/558,257

NUMBER OF SEQ ID NOS: 15

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 8

LENGTH: 9726

TYPE: DNA

ORGANISM: Fusobacterium necrophorum

US-09-841-786-8

Query Match 21.3%; Score 31.8; DB 10; Length 9726;

Best Local Similarity 56.1%; Pred. No. 4.8;

Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 42 AGAGCGAAGAACTTCAGACCTGAGTCAATCTATCGGAAATCCAGAAATTAAG 101

DB 2082 AGAAGAAAGAACTTATGGAATGAGAGAGTTTATGGAAGCTATTAATATATGC 2141

QY 102 TGAAGTAAATATGTCCCAAGCTGAACAGTACTAGAGAGAAAA 148

DB 2142 TTCTGTGACAAATGCCGATCATGCTGAATGGTATCGAAGAAAA 2188

RESULT 8

US-09-841-786-15

Sequence 15, Application US/09841786

Patent No. US20020054883A1

GENERAL INFORMATION:

APPLICANT: NAGARAJA, T. G.

APPLICANT: STEWART, GEORGE C.

APPLICANT: NARAYANAN, SANJEEV K.

TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN

FILE REFERENCE: 30296

CURRENT APPLICATION NUMBER: US/09/841,786

PRIOR FILING DATE: 2001-04-24

PRIOR APPLICATION NUMBER: 09/558,257

NUMBER OF SEQ ID NOS: 15

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 15

LENGTH: 11130

TYPE: DNA

ORGANISM: Fusobacterium necrophorum

US-09-841-786-15

Query Match 21.3%; Score 31.8; DB 10; Length 11130;

Best Local Similarity 56.1%; Pred. No. 5.1;

Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 42 AGAGCGAAGAACTTCAGACCTGAGTCAATCTATCGGAAATCCAGAAATTAAG 101

DB 3115 AGAAGAAAGAACTTATGGAATGAGAGAGTTTATGGAAGCTATTAATATATGC 3174

QY 102 TGAAGTAAATATGTCCCAAGCTGAACAGTACTAGAGAGAAAA 148

DB 3175 TTCTGTGACAAATGCCGATCATGCTGAATGGTATCGAAGAAAA 3221

RESULT 9

US-09-815-242-6627

Sequence 6627, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: HASELBECK, ROBERT

APPLICANT: OHLESEN, KARI L.

APPLICANT: ZYSKIND, JUDITH W.

APPLICANT: WALL, DANIEL

APPLICANT: TRAWICK, JOHN D.

APPLICANT: CARR, GRANT J.

APPLICANT: YAMAMOTO, ROBERT T.

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA, 011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6627
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(873)
US-09-815-242-6627

```

```

Query Match          20.1%; Score 30; DB 10; Length 873;
Best Local Similarity 57.4%; Pred. No. 6.9;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

QY 46 GCGAGAACTTCCAGAGCTGCGATCTATCGGAAATCCAGATATTAAAGTAA 105
DB 254 GCGAGAAATTTCTGTCAGCAGCAAAATTTAGATGAAATGAGCTATGAAATTCGA 313
QY 106 CTAAATATGTCCCAAGCTGAACTAGTAGG 139
DB 314 TTAAAGCATTTGATTAAGCTTAGAAAAATTAGG 347

```

```

RESULT 10
US-09-764-887-338
; Sequence 338, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, proteins, and antibodies
; FILE REFERENCE: PA113
; CURRENT APPLICATION NUMBER: US/09/764,887
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 338
; LENGTH: 32190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-887-338

```

```

Query Match          20.0%; Score 29.8; DB 10; Length 32190;
Best Local Similarity 52.9%; Pred. No. 32;
Matches 64; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```

```

QY 2 GTGTTCTCCACCTCGAGATGAGAGAAAGCAATTCAGAGCAAGAACTTCAG 61
DB 13182 GTATCTCTACTACTTGGCCCTCCCAAAATATCTGGGATTTATGGGTGAGCCACTGTACT 13241
QY 62 GACCTGAGTCATCTATCGGAAATCCAGATATTAAAGTAACTAAATATGTCCCA 121
DB 13242 GGCTGAGATATTTTAAATAATTTCTTAAACTTAATACACACAAAAAACCCTAA 13301
QY 122 A 122
DB 13302 A 13302

```

RESULT 11

```

US-09-969-347-302/C
; Sequence 302, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
; APPLICANT: Ederer, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-69
; CURRENT APPLICATION NUMBER: US/09/969,347
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,598
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,604
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 318
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 302
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-347-302

```

```

Query Match          19.9%; Score 29.6; DB 10; Length 425;
Best Local Similarity 59.5%; Pred. No. 7;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

QY 56 TTCCAGAGCTGCGATCTATCGGAAATCCAGATATTAAAGTAACTTAATATG 115
DB 290 TTCCAGTAAATGAGTTTCTGTAAGAAATCCAAAAAGCAAAATTAATGATTTT 231
QY 116 TCCCAAGCTGAACTAGTAGG 139
DB 230 TTCTGCTGCTGAACTAGTAGG 207

```

```

RESULT 12
US-09-803-719-278/C
; Sequence 278, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Gleise, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassem, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-crain, Birgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624,002
; CURRENT APPLICATION NUMBER: US/09/803,719
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 278
; LENGTH: 329
; TYPE: DNA

```

PRIOR FILING DATE: 2000-02-04

APPLICANT: Botstein, David

APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerltzen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C11
CURRENT APPLICATION NUMBER: US/09/978,295A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920

PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29

;; PRIOR APPLICATION NUMBER: 60/083559
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083500
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: 60/083742
;; PRIOR FILING DATE: 1998-04-30
;; PRIOR APPLICATION NUMBER: 60/084366
;; PRIOR FILING DATE: 1998-05-05
;; PRIOR APPLICATION NUMBER: 60/084414
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084441
;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084639
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084640
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084598
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084627
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/084643
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085582
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085700
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085689
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085580
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085573
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697

Query Match 19.64; Score 29.2; DB 9; Length 2558;
Best Local Similarity 50.74; Pred. No. 19;
Matches 70; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 6 TCCCTCCACCTCGATGAGAGAAAGCAATTCAGAGAGCAAGAACTTCAGAGACC 65
DB 876 TCGTACTACTTGTCTCTCGGGTGAAGTCTTATCCAGACGGTTGGAATCTTCTCGAGG 935
QY 66 TGCAGTCATCTATCCGAATCCGAATATTAAAGTGAATCTAAATATGTCCCAAGC 125
DB 936 TGGGTCTCAGCGTGAATAATCTTAAATCTGAATGTCAGAGACCTCTCAGACCAG 995
QY 126 TGAACAGTAGTAGAAGA 143
DB 996 TTACCCAGCAAAATGAATA 1013

Search completed: April 16, 2003, 01:17:33
Job time : 200.888 secs

GenCore version 5.1.4.D5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:26:41; Search time 39.7606 Seconds

(without alignments)
8439.207 Million cell updates/sec

Title: US-09-647-019-9

Perfect score: 149

Sequence: 1 ggtgtccctccaccctcgga.....cagtagragsagaagaaag 149

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: N_Geneseq_101002.*
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149	100.0	290	24	AAD27226
2	149	100.0	886	24	AAD27216
3	149	100.0	887	20	AAZ90904
4	112.8	75.7	587	24	ABK64187
5	112.8	75.7	587	24	ABK64590
6	96.2	64.6	778	20	AAZ90903
7	49	32.9	428	21	AAZ01483
8	47.4	31.8	65	24	ABN58272
9	33.4	22.4	11049	24	ABL92218

C 10	33.4	22.4	11049	24	ABL49321
C 11	33.4	22.4	11049	24	ABL32668
C 12	32.8	22.0	640681	20	ABA92787
C 13	32.6	21.9	910715	20	AAZ0248
C 14	32	21.5	10589	23	ABZ8294
C 15	31.8	21.3	615	23	ABV34906
C 16	31.8	21.3	615	23	ABV43755
C 17	31.8	21.3	796	22	AA196124
C 18	31.8	21.3	810	23	AA196124
C 19	31.8	21.3	2780	24	AA167642
C 20	31.8	21.3	9726	24	AA167642
C 21	31.8	21.3	11130	24	AA167647
C 22	31.2	20.9	730	24	AA562053
C 23	30.6	20.5	198	21	AAZ00359
C 24	30.6	20.5	451	21	AAZ00359
C 25	30.4	20.4	2261	20	AAZ03588
C 26	30.2	20.3	1515	21	AAZ03588
C 27	30.2	20.3	3415	23	AAZ03588
C 28	30.2	20.3	17934	24	ABL04838
C 29	30.2	20.3	17934	24	ABL33718
C 30	30	20.1	377	22	AA181299
C 31	30	20.1	810	23	AA566690
C 32	30	20.1	873	23	AA575823
C 33	30	20.1	9760	24	AA575823
C 34	30	20.1	9760	24	AA575823
C 35	30	20.1	9760	24	AA575823
C 36	30	20.1	9760	24	AA575823
C 37	29.8	20.0	5236	24	ABL32350
C 38	29.8	20.0	5236	24	ABL32350
C 39	29.8	20.0	5474	24	ABL32350
C 40	29.8	20.0	32190	24	ABL32350
C 41	29.8	20.0	32190	24	ABL32350
C 42	29.8	20.0	32190	24	ABL32350
C 43	29.8	20.0	40324	24	ABN50217
C 44	29.8	20.0	121724	24	ABO67149
C 45	29.6	19.9	425	24	ABO88143

ALIGNMENTS

RESULT 1	
ID	AAD27226 standard; CDNA; 290 BP.
XX	AAZ90904
AC	AAZ90904
DT	09-APR-2002 (first entry)
XX	
XX	
DE	Human 66214 CDNA clone.
XX	
XX	
KW	Human; congestive heart failure; dilative cardiomyopathy; sudden death;
KW	hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
KW	heart muscle disease; conduction disorder; coronary heart disease;
KW	systemic arterial hypertension; pulmonary hypertension; endocarditis;
KW	pulmonary heart disease; valvular heart disease; pericardial disease;
KW	congenital heart disease; gene therapy; syncope; transgenic animal;
KW	clone 66214; ss.
XX	
XX	
XX	Homo sapiens.
XX	
XX	MO200192567-A2.
XX	
PD	06-DEC-2001.
XX	
XX	30-MAY-2001; 2001MO-EP06165.
XX	
XX	30-MAY-2000; 2000US-207400P.
XX	
PA	(MED1-) MEDIGENE AG.
XX	
XX	
PI	Bunk D, Reuner B, Beck J, Henkel T,

Human polynucleotide
Human immune system
Bacteria sp. genome
Borrelia burgdorferi
Drosophila melanogaster
Human prostate exp
Human prostate exp
Human neuroblastoma
Human prostate exp
F. necrophorum leu
F. necrophorum leu
Porcine muscular s
Human secreted pro
Human secreted pro
Human breast tumor
Plasmodium falciparum
Drosophila melanogaster
Human immune system
Human polynucleotide
DNA encoding novel
DNA encoding novel
Enterococcus faecalis
Chemically treated
Human gene regulat
Signal transductio
Arabidopsis thaliana
Human secreted pro
Human immune system
Human immune system
Human digestive system
Human liver antigen
Human angiogenesis
Human osteoblast d
Pancreatic cancer re

Chisel gene; Csl; EF-Hand protein super family; muscle development;
heart/skeletal muscle cell development; signaling pathway; regulation,
Xq21.3-q22; adaptive process; muscle homeostasis; skeletal myopathy;

KM detection; diagnosis; prophylaxis; treatment; cardiac hypertrophy;
 KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KM Becker's myotonic dystrophy; heart failure; differentiation; exon;
 KM gene therapy; transgenic animal; drug screening; ss.
 OS
 XX Homo sapiens.

Key Location/Qualifiers
 exon 1..172
 /tag= a
 /label= Exon_1
 /note= "Corresponds to residues 19497-19327 of human cosmid clone U228D4"
 173..229
 /tag= b
 /label= Exon_2
 /note= "Corresponds to residues 15687-15631 of human cosmid clone U228D4"
 185..451
 /tag= c
 /product= "Chisel (Csl) protein"
 /note= "Expressed predominantly in heart muscles"
 230..316
 /tag= d
 /label= Exon_3
 /note= "Corresponds to residues 5220-5134 of human cosmid clone U228D4"
 317..465
 /tag= e
 /label= Exon_4
 /note= "Corresponds to residues 35384-35236 of human cosmid clone U112E8"
 466..887
 /tag= f
 /label= Exon_5
 /note= "Corresponds to residues 4101-3680 of human cosmid clone U112E8"
 exon
 W09950410-A1.

07-OCT-1999.
 26-MAR-1999; 99MO-AU00220.
 27-MAR-1998; 98AU-0002634.
 (CHAN-) CHANG CARDIAC RES INST VICTOR.
 (GENO-) GEN HOSPITAL CORP.
 (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 PI Harvey RP, Muscareo A, Palmer SJ, Rosenthal NA,
 WPI; 1999-610852/52.
 DR P-PSDB; AAY28651.
 XX
 PT Isolated nucleic acids encoding chisel, used to develop products for
 PT treating cardiomyopathy, cardiac hypertrophy, heart failure and
 PT muscular myopathies -
 XX
 PS Claim 8; Page 149-150; 157pp; English.

CC The present sequence is the cDNA encoding the human chisel gene (Csl)
 CC that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.
 CC It is a member of the EF-Hand protein super family and is involved in
 CC signalling pathways. It is predominantly expressed in heart and skeletal
 CC muscles and is activated after the differentiation of cells. Csl
 CC functions in regulation aspects of differentiation or adaptive processes
 CC that maintain muscle homeostasis. This sequence can be used in the
 CC detection, diagnosis, prophylactic and therapeutic treatment of diseases
 CC such as those involving aberrant muscle cell development and functional
 CC activity. It is also used in the treatment of muscular and myotonic
 CC dystrophies, skeletal myopathies such as Duchenne muscular dystrophy and
 CC Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,

CC myocarditis, myofiber atrophy, etc. The Csl gene sequence can also be
 CC used in gene therapy, for the production of transgenic animals and for
 CC drug screening.
 XX
 SQ Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;

Query Match 100.0%; Score 149; DB 20; Length 887;
 Best Local Similarity 100.0%; Pred. No. 1.5e-36;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGTTCCTCCCACTCGATGAGAGAGAGCCCAATTCAGAGCGAAGAACTTCCA 60
 Db 317 GGTTCCTCCCACTCGATGAGAGAGAGCCCAATTCAGAGCGAAGAACTTCCA 376
 Qy 61 GAACTGCACTCAATCTATCGAATTCAGAAATATTAAGTGAATAATATGTCGCC 120
 Db 377 GAACTGCACTCAATCTATCGAATTCAGAAATATTAAGTGAATAATATGTCGCC 436
 Qy 121 AAGCTGAACAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 149
 Db 437 AAGCTGAACAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465

RESULT 4
 ABK64187/c
 ID ABK64187 standard; DNA; 587 BP.
 XX
 AC ABK64187;
 XX

DT 18-JUN-2002 (first entry)

DE Human benign prostatic hyperplasia gene #82.

XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.

OS Homo sapiens.

XX W0200212440-A2.

PD 14-FEB-2002.

PF 07-AUG-2001; 2001WO-US24708.

PR 07-AUG-2000; 2000US-223323P.

PR 05-JUN-2001; 2001US-0873319.

PA (GENE-) GENE LOGIC INC.

PA (NLSB) JAPAN TOBACCO INC.

PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;

WPI; 2002-257476/30.

PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
 PT detecting expression levels of one or more genes in prostate cells from
 PT patient that are differentially regulated compared to normal prostate
 PT cells -
 XX

PS Disclosure; Page 104; 444pp; English.

CC The invention relates to a method of diagnosing (I) the onset or
 CC progression of benign prostatic hyperplasia (BPH), or screening (II) for
 CC or identifying an agent that modulates the onset or progression of BPH.
 CC The method is based on changes in gene expression in BPH tissue isolated
 CC from patients exhibiting different clinical states of prostate
 CC hyperplasia as compared to normal prostate tissue. (I) comprises
 CC detecting the expression levels of one or more genes in prostate cells
 CC from the subject that are differentially regulated compared to normal
 CC prostate cells. (II) comprises preparing a first gene expression profile
 CC of BPH cells or BPH-like cell population, exposing the cells to the
 CC agent, preparing a second gene expression profile of the cells to the
 CC cells, and comparing the first and second gene expression profiles.
 CC (I) is useful for diagnosing the onset or progression of BPH. (II) is

PR 27-SEP-2000; 2000US-235863P

RESULT 6

ID ABNS8272 standard; DNA; 65 BP.
 XX
 AC ABNS8272;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:31020.
 XX
 KM Human; mouse; rat; splice transcript; detection; RNA transcript;
 XX splice variant; transcriptome; oligonucleotide library; ss.
 XX
 OS Mus musculus.
 XX
 PN WO200210449-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 20-JUL-2001; 2001WO-1B01903.
 XX
 PR 28-JUL-2000; 2000US-221607P.
 XX
 PR 02-MAY-2001; 2001US-287724P.
 XX
 PA (COMP.) COMPUGEN INC.
 XX
 PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
 XX
 DR WPI; 2002-257383/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes -
 XX
 PS Example 1; SEQ ID 31020; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterizing the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABNS27253 to ABNS9589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIGO
 CC at ftp.wigo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 65 BP; 24 A; 10 C; 13 G; 18 T; 0 other;
 XX
 Query Match 31.8%; Score 47.4; DB 24; Length 65;
 Best Local Similarity 83.1%; Pred. No. 2.5e-05;
 Matches 54; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 XX
 QY 48 GAAGAACTTCGAGACTGCTATCGAAATCCGAAATATTAAGTAACT 107
 DB 1 GAAGAAATTTCCGAGACTGCTATCGAAATCCGAAATATTAAGTAACT 60
 QY 108 AAAAT 112
 DB 61 GAAAT 65

RESULT 9
 ID ABL92218/c
 XX ABL92218 standard; DNA; 11049 BP.
 XX
 AC ABL92218;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Chemically treated DNA repair gene fragment#14.
 XX
 KM DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L13;
 KM PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;
 KM DDIRL; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;
 KM Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;
 KM immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;
 KM cancer; ds.
 XX
 OS Unidentified.
 XX
 PN WO200181622-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03972.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 XX
 PR 07-APR-2000; 2000DE-1019173.
 XX
 PR 30-JUN-2000; 2000DE-1035529.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-034446/04.
 XX
 PT New nucleic acid derived from genes associated with DNA repair, useful
 PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 27; 25pp + sequence listing; English.
 XX
 CC The invention relates to nucleic acids containing a sequence of at least
 CC 18 nucleotides of chemically treated DNA of genes associated with DNA
 CC repair, and their complements. The invention also relates to nucleic
 CC acids comprising at least 18 base pairs of the chemically pretreated DNA
 CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
 CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
 CC RFC4, DDIRL, FANCB, or XRCC8. Nucleic acids of the invention and related
 CC oligomers, are useful for diagnosis of diseases associated with gene
 CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
 CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
 CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours
 CC and cancer, particularly by determining status of cytosine methylation
 CC and/or by detecting single-nucleotide polymorphisms. Determination of
 CC individual methylation patterns may allow development of individualised
 CC therapies. The sequences given in records ABL92192-ABL92335 represent
 CC chemically pre-treated DNA fragments from genes associated with DNA
 CC repair, and their complements.
 CC Note: The sequence data for this patent is not represented in the
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX
 SQ Sequence 11049 BP; 3032 A; 198 C; 2438 G; 5381 T; 0 other;
 XX
 Query Match 22.4%; Score 33.4; DB 24; Length 11049;
 Best Local Similarity 54.5%; Pred. No. 2.6;
 Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
 QY 26 AGAAGAGCAATTCGAGAGCAAGAACTTCGAGAGCTGCAATCTATCGAAA 85


```

QY 26 AGAAGAGCCCAATTCAGAGCGAAGAACTTCAGACCTCGATCATCTATCGGAA 85
DB 5771 AAAAAAAAAAAGCAAGCCGCTCATCACCGATTCGCAATCACTACACGAA 5712
QY 86 TCAGAAATATTAAGAGCACTAAATATGTCCCAAGCTGAAGTAGTAGAAGAA 145
DB 5711 TCAGAAATATTAAGAGCACTAAATATGTCCCAAGCTGAAGTAGTAGAAGAA 5652
QY 146 AAA 148
DB 5651 AAA 5649

RESULT 12
ABA92787
ID ABA92787 standard; DNA; 640681 BP.
XX ABA92787;
XX 27-MAR-2002 (first entry)
XX Buchnera sp. genomic DNA SEQ ID NO:1.
XX Buchnera sp. genomic DNA SEQ ID NO:1.
XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
XX circular; ds.
XX Buchnera sp.
XX Buchnera sp.
XX JP2001292771-A.
XX 23-OCT-2001.
XX 07-APR-2000; 2000JP-0107160.
XX 07-APR-2000; 2000JP-0107160.
XX (RIKA) RIKAGAKU KENKYUSHO.
XX WPI; 2002-126043/17.
XX A genomic DNA of cockroach-symbiotic bacterium
XX Claim 1; Page 16-230; 237pp; Japanese.
XX The present invention describes a gene (I) derived from Buchnera sp.
XX containing the DNA (a) or (b), (a) has a fully defined base pair
XX sequence selected from a table of sequences found in the Buchnera sp.
XX genomic DNA of ABA92787 given in the specification or is a DNA selected
XX from complementary DNA sequences, and (b) is a DNA which hybridizes with
XX the DNA (a) and encodes a protein. Also described are: (1) a recombinant
XX vector (II) containing (I); (2) a transformant (III) containing (II);
XX (3) a genomic DNA of Buchnera sp. containing the sequence given in
XX ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
XX (d) (c) is a DNA containing a fully defined sequence given in ABA92788
XX or ABA92789 and (d) is a plasmid which hybridizes with a DNA; and (5) a
XX method for the preparation of a protein in which (III) is cultured and
XX the expression protein of the objective protein is collected from the
XX resultant culture. The DNA is useful for developing agricultural
XX chemicals for exterminating cockroaches. The present sequence represents
XX the specifically claimed Buchnera sp. genomic DNA sequence, from the
XX present invention.
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
SQ

Query Match 22.0%; Score 32.8; DB 24; Length 640681;
Best Local Similarity 58.0%; Pred. No. 14;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 49 AAGAACTTCAGAGCACTGATCATCTATCGCAATTCAGAAATATTAAGTAGAAGTA 108
DB 520435 AATTAAGTAGAAGCACTGATCTTTTATTCAGCAACCACTATTAATATGAGACTA 520494
QY 109 AATATGTGTCCTCAAGCTGAACAGTAGTAGAAGAAAAA 148

```

```

DB 520495 ATGATTTTATTTATCTCAGCGTCATTAATAAAAAATAAA 520534

RESULT 13
AAX20248/c
ID AAX20248 standard; DNA; 910715 BP.
XX AAX20248;
XX 04-MAY-1999 (first entry)
XX Borrelia burgdorferi polynucleotide sequence #1.
XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
XX epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX infection; diagnosis; characterisation; detection; ds.
XX Borrelia burgdorferi.
XX WO9858943-A1.
XX 30-DEC-1998.
XX 18-JUN-1998; 98WO-US12764.
XX 03-SEP-1997; 97US-0057483.
XX 20-JUN-1997; 97US-0050359.
XX 22-JUL-1997; 97US-0053344.
XX 22-JUL-1997; 97US-0053377.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (MEDI-) MEDIMADNE INC.
XX Claydon R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX White OR;
XX WPI; 1999-081217/07.
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
XX products for the detection, diagnosis, characterisation, prevention
XX and therapy of infections, particularly Lyme disease
XX Claim 1; Page 157-671; 1128pp; English.
XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from
XX Borrelia burgdorferi (Bb). Products derived from Bb can be used for
XX the detection, diagnosis, characterisation, prevention and therapy of
XX Bb infections, e.g. Lyme disease. They can also be used for the
XX production of biosynthetic products, e.g. enzymes. Borrelia belongs
XX to a family of motile, spiral-shaped bacteria called Spirochetes.
XX Spirochetes are pathogenic in humans and Borrelia causes epidemic and
XX endemic relapsing fever, and Lyme borreliosis, more commonly known as
XX Lyme disease.
XX Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other;
SQ

Query Match 21.9%; Score 32.6; DB 20; Length 910715;
Best Local Similarity 69.8%; Pred. No. 18;
Matches 44; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 81 GGAAATCCAGATATTTAAAGTGAACCTAAATATGTCCTCCCAAGCTGAACAGTAGTAGA 140
DB 706680 GGAAATCCAGATATTTAAATTAAGCAATTAATCTTATCAATTAAGTAGTAGA 706621
QY 141 AGA 143
DB 706620 AAA 706618

RESULT 14
ABL28294
ID ABL28294 standard; DNA; 10589 BP.

```



```

XX AC ABL28294;
XX XX
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 36355.
XX XX
XX KM Drosophila; developmental biology; cell signalling; insecticide;
XX XX pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX XX
XX PN MO200171042-A2.
XX XX
XX PD 27-SEP-2001.
XX XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX XX
XX PR 23-MAR-2000; 2000US-191637P.
XX XX
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX PA (PEKE ) PB CORP NY.
XX XX
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX XX
XX DR WPI; 2001-656860/75.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX XX interactions -
XX XX
XX PS Claim 1; SEQ ID NO 36355; 21pp + Sequence listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX XX useful in developmental biology and in elucidating cell signalling and
XX XX cell-cell interactions in higher eukaryotes for the development of
XX XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX XX sequences (AB101840-AB16175) and the encoded proteins
XX XX (AAB5737-AB572072).
XX XX
XX CC The sequence data for this patent did not form part of the printed
XX XX specification, but was obtained in electronic format directly from WIPO
XX XX at ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 10589 BP; 3083 A; 2231 C; 2183 G; 3092 T; 0 other;

Query Match 21.5%; Score 32; DB 23; Length 10589;
Best Local Similarity 60.2%; Pred. No. 7.1;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 47 CGAAGAACTCCAGACCTGCACTCAATCTATCGAAATCCAGAAATATTAAGTGAC 106
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1744 CGTAAACCTTACAGCGCTGCACTTATGAAAAGTTATGCGAAATATTAACAAAG 1803
OY 107 TAAATATGTGCTCCCAAGCTGAACAGTA 134
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1804 TAACTGTTCACGAAACAGTACAAAGTA 1831

RESULT 15
ID ABV34906 standard; cDNA; 615 BP.
XX ABV34906;
XX AC
XX AC ABV34906;
XX XX
XX DT 16-SEP-2002 (first entry)
XX XX
XX DE Human prostate expression marker cDNA 34897.
XX XX
XX KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX XX pharmacogenomic marker; gene; ss.

```

```

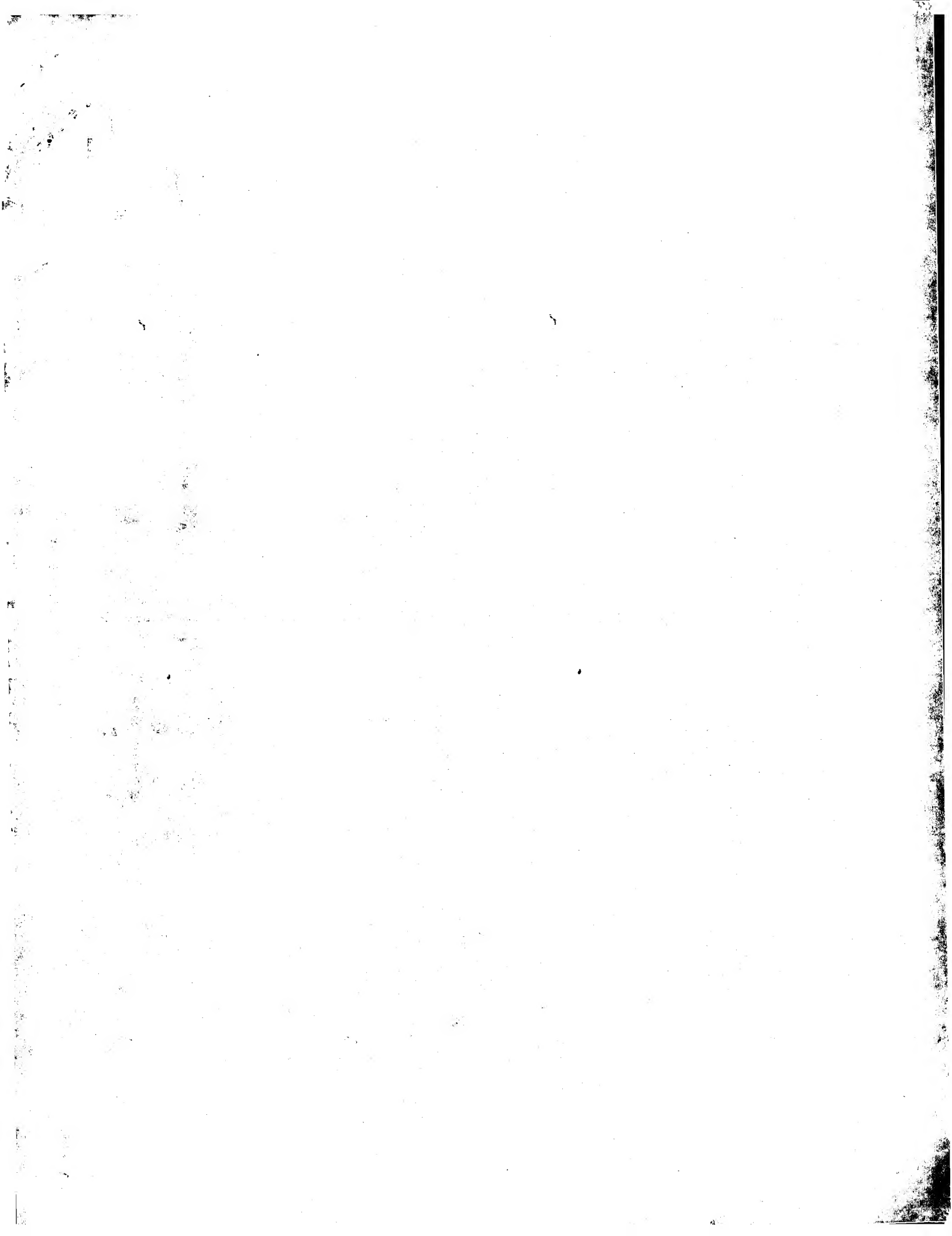
OS Homo sapiens.
XX XX
XX PN WO200160860-A2.
XX XX
XX PD 23-AUG-2001.
XX XX
XX PF 20-FEB-2001; 2001WO-US05171.
XX XX
XX PR 17-FEB-2000; 2000US-183319P.
XX XX
XX PR 16-MAR-2000; 2000US-189862P.
XX XX
XX PR 25-MAY-2000; 2000US-207454P.
XX XX
XX PR 09-JUN-2000; 2000US-211314P.
XX XX
XX PR 18-JUL-2000; 2000US-219007P.
XX XX
XX PR 13-DEC-2000; 2000US-255281P.
XX XX
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX XX
XX DR WPI; 2001-662795/76.
XX XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX XX prostate cells and correlating with presence of prostate cancer, useful
XX XX for detecting presence of prostate cancer, stage of prostate cancer -
XX XX
XX PS Claim 1; Page 7298; 11750pp; English.
XX XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX XX specification or its complement. (I) is useful for:
XX XX
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX XX
XX CC (b) monitoring the efficacy of a test compound to inhibit prostate
XX XX cancer in a patient;
XX XX
XX CC (c) assessing the efficacy of a therapy for inhibiting prostate cancer
XX XX in a patient;
XX XX
XX CC (d) selecting a composition for inhibiting prostate cancer in a patient;
XX XX
XX CC (e) assessing the prostate cell carcinogenic potential of a compound;
XX XX
XX CC (f) determining whether prostate cancer has metastasized in a patient;
XX XX
XX CC (g) assessing the aggressiveness or indolence of prostate cancer in a
XX XX patient;
XX XX
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX XX
XX SQ Sequence 615 BP; 208 A; 131 C; 140 G; 136 T; 0 other;

Query Match 21.3%; Score 31.8; DB 23; Length 615;
Best Local Similarity 51.0%; Pred. No. 3.4;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

OY 2 GTGTCCTCCCACTCCGATGAGAGAAAGCCAAATTCAGAGCGAAGAACTTCAG 61
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 GTGATCCGCCACCTCGGCTCCCAAGTGTGGATTACAGCGCGACACCATACCC 407
OY 62 GACCTGACATCATCTATCGAAATCCAGAAATATTAAGTAAGTAATATGTCCCA 121
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 408 GGCCCTTATACATTAAGAGATTTTATAAAGATTAATTAATTCACATATAGAGA 467
OY 122 AAGCTGAACATGATGAGAGAAAAA 148
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 468 AAACGTGAGAGAGAAAGATGACAA 494

Search completed: April 15, 2003, 19:04:58
Job time : 415.761 secs

```



GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model1

Run on: April 15, 2003, 16:47:06 Search time 337.585 Seconds

(without alignment) 12645.099 Million cell updates/sec

Title: US-09-647-019-9

Perfect score: 149

Sequence: 1 99tgcctccctccacctcgga.....cagtagtagaagaaagaaag 149

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

```

1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_rtr:*
12: gb_by:*
13: gb_un:*
14: gb_vl:*
15: gb_vl:*
16: gb_vl:*
17: em_fun:*
18: em_fun:*
19: em_in:*
20: em_mu:*
21: em_mu:*
22: em_mu:*
23: em_mu:*
24: em_mu:*
25: em_mu:*
26: em_mu:*
27: em_mu:*
28: em_mu:*
29: em_mu:*
30: em_mu:*
31: em_mu:*
32: em_mu:*
33: em_mu:*
34: em_mu:*
35: em_mu:*
36: em_mu:*
37: em_mu:*
38: em_mu:*
39: em_mu:*
40: em_mu:*
41: em_mu:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	149	100.0	240	6	AX322783	AX322783 Sequence
2	149	100.0	835	9	BC005948	BC005948 Homo sapi
3	149	100.0	885	9	HS4250584	HS4250584 Homo sapi
4	149	100.0	886	6	AX322774	AX322774 Sequence
5	149	100.0	886	9	AF129505	AF129505 Homo sapi
6	149	100.0	36503	9	U73508	U73508 Homo sapi
7	149	100.0	150319	2	AL772370	AL772370 Homo sapi
8	112.8	75.7	587	6	AX32418	AX32418 Sequence
9	96.2	64.6	787	10	AY026524	AY026524 Mus muscu
10	96.2	64.6	936	10	AF364070	AF364070 Mus muscu
11	96.2	64.6	943	10	MMU245772	MMU245772 Mus muscu
12	96.2	64.6	228031	2	AF364071	AF364071 Mus muscu
13	85.8	57.6	892	5	AF343894	AF343894 Xenopus laevis
14	73.2	49.1	923	5	AF343894	AF343894 Xenopus laevis
15	36.6	24.6	42446	8	ATFL73519	ATFL73519 Arabidops
16	36.6	24.6	91228	8	ATFL73519	ATFL73519 Arabidops
17	36.6	24.6	198777	8	ATFL73519	ATFL73519 Arabidops
18	36.6	24.6	175166	2	AL840631	AL840631 Arabidops
19	35.6	23.9	293431	2	PFMAL13P4	PFMAL13P4 Arabidops
20	35.2	23.6	171004	2	AC006286	AC006286 Homo sapi
21	35.2	23.6	171004	2	AC006286	AC006286 Homo sapi
22	35.2	23.6	171004	2	AC006286	AC006286 Homo sapi
23	35.2	23.6	171004	2	AC006286	AC006286 Homo sapi
24	34.8	23.4	164552	2	AL357055	AL357055 Homo sapi
25	34.8	23.4	164552	2	AL357055	AL357055 Homo sapi
26	34.4	23.1	127507	2	AP005311	AP005311 Oryza sat
27	34.4	23.1	127507	2	AP005311	AP005311 Oryza sat
28	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
29	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
30	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
31	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
32	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
33	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
34	34.2	23.0	143494	9	AL391497	AL391497 Human DNA
35	33.8	22.7	176368	9	CNS01RGG	CNS01RGG Homo sapi
36	33.8	22.7	176368	9	CNS01RGG	CNS01RGG Homo sapi
37	33.8	22.7	176368	9	CNS01RGG	CNS01RGG Homo sapi
38	33.8	22.7	176368	9	CNS01RGG	CNS01RGG Homo sapi
39	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
40	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
41	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
42	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
43	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
44	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi
45	33.6	22.6	102923	9	AC104803	AC104803 Homo sapi

ALIGNMENTS

RESULT 1
AX322783
LOCUS
DEFINITION
AX322783
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
Bunk,D., Reuner,B., Beck,J. and Henkel,T.
Novel target genes for diseases of the heart
Patent: WO 012567-A 27 06-DEC-2001;
Medigene AG (DE)

AX322783
Sequence 27 from Patent WO0192567.
AX322783.1 GI:18093762
290 bp
DNA
linear
PAT 07-JAN-2002

FEATURES
source
Location/Qualifiers
1. .290
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 103 a 50 c 69 g 68 t

ORIGIN

Query Match 100.0%; Score 149; DB 6; Length 290;
Best Local Similarity 100.0%; Pred. No. 8, 9e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 60
|||
DB 19 GGTCTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 78
|||
QY 61 GGACCTGAGTCAATCTATCCGAAATCCAGATTTAAAGTGAATTAATGTCTCC 120
|||
DB 79 GGACCTGAGTCAATCTATCCGAAATCCAGATTTAAAGTGAATTAATGTCTCC 138
|||
QY 121 AAAGCTGACGATGATGAGAGAGAGAG 149
|||
DB 139 AAAGCTGACGATGATGAGAGAGAGAG 167
|||

RESULT 2
LOCUS BC005948 835 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, small muscle protein, X-linked, clone MGC:14584
IMAGE:4246501, mRNA, complete cds.
ACCESSION BC005948
VERSION BC005948.1 GI:13543590
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 835)
Strausberg, R.
Direct Submission
Submitted (02-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-rc@mail.nih.gov
Tissue Procurement: CLOMTECH
CDNA Library Preparation: CLOMTECH Laboratories, Inc.
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdepxall.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>
Series: IRAL Plate: 21 Row: a Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6625646.
Location/Qualifiers
1. .835
/organism="Homo sapiens"
/db_xref="taxon:32644"
/db_xref="taxon:9606"
/clone="MGC:14584 IMAGE:4246501"
/tissue type="Skeletal Muscle"
/clone_lib="NIH_MGC_81"
/lab host="DH10B"
/note="Vector: pDNR-LIB"
113 .379
/codon_start=1

FEATURES
source
Location/Qualifiers
1. .835
/organism="Homo sapiens"
/db_xref="taxon:32644"
/db_xref="taxon:9606"
/clone="MGC:14584 IMAGE:4246501"
/tissue type="Skeletal Muscle"
/clone_lib="NIH_MGC_81"
/lab host="DH10B"
/note="Vector: pDNR-LIB"
113 .379
/codon_start=1

BASE COUNT 282 a 155 c 171 g 227 t

ORIGIN

Query Match 100.0%; Score 149; DB 9; Length 835;
Best Local Similarity 100.0%; Pred. No. 8, 8e-33;
Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTCTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 60
|||
DB 245 GGTCTCTCCACCTCGATGAGAGAGAGCAATTCAGAGCGAAGAACTTCCA 304
|||
QY 61 GGACCTGAGTCAATCTATCCGAAATCCAGATTTAAAGTGAATTAATGTCTCC 120
|||
DB 305 GGACCTGAGTCAATCTATCCGAAATCCAGATTTAAAGTGAATTAATGTCTCC 364
|||
QY 121 AAAGCTGACGATGATGAGAGAGAGAG 149
|||
DB 365 AAAGCTGACGATGATGAGAGAGAGAG 393
|||

RESULT 3
LOCUS HSA250584 885 bp mRNA linear PRI 12-APR-2001
DEFINITION Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
protein (Strmx gene).
ACCESSION AJ250584
VERSION AJ250584.1 GI:10178976
KEYWORDS Strmx gene; stretch responsive muscle (X-chromosome).
SOURCE human.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 885)
Kemp, T.J., Sadusky, T.J., Simon, M., Brown, R., Eastwood, M.,
Sassoon, D.A. and Coulton, G.R.
Identification of a novel stretch-responsive skeletal muscle gene
(Smx)
Genomics 72 (3), 260-271 (2001)
JOURNAL MEDLINE 21295047
PUBMED 11401441
REFERENCE 2 (bases 1 to 885)
AUTHORS Kemp, T.J.
TITLE Direct Submission
Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial
College of Science Technology and Medicine, SAF Bldg, Level 2,
Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
Related sequences: AJ245772, U73508 to U73509.
Location/Qualifiers
1. .885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue type="Skeletal muscle"
1. .885
/gene="Strmx"
1. .183
/gene="Strmx"
184 .450
/gene="Strmx"
/gene="Strmx"
/codon_start=1
/product="stretch responsive muscle (X-chromosome)"
/protein_id="CAC08492.1"
/db_xref="GI:10178976"
/translation="NMMSKQPVSNVRAIOANINIPWGAFFPGAGPPRRKCTPEVEE
GVPTSPDEKKPIPGAKKLPGPAVNLSIIONIKSEIKYVPAEQ"
451 .885
/gene="Strmx"
857 .862
/gene="Strmx"
/gene="Strmx"

BASE COUNT	278 a	172 c	191 g	244 t
ORIGIN				
Query Match	100.0%;	Score 149;	DB 9;	Length 885;
Best Local Similarity	100.0%;	Pred. No. 8.8e-33;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	1	GGGTTCCTCCCACTTCGGATGAGAGAAAGCCCAATTCGAGAGCGAAGAACTTCCA	60	
Db	316	GGGTTCCTCCCACTTCGGATGAGAGAAAGCCCAATTCGAGAGCGAAGAACTTCCA	375	
Qy	61	GGACCTGACAGTCAATCTATCGGAAATCCGAAATATTAAAGTAACTTAATATGTCCCC	120	
Db	376	GGACCTGACAGTCAATCTATCGGAAATCCGAAATATTAAAGTAACTTAATATGTCCCC	435	
Qy	121	AAAGCTGAACAGTAGTAGAGAAAGAAAG	149	
Db	436	AAAGCTGAACAGTAGTAGAGAAAGAAAG	464	
RESULT 4				
AX322774				
LOCUS	AX322774	886 bp	DNA	linear
DEFINITION	Sequence 18 from Patent WO0192567.			
ACCESSION	AX322774			
VERSION	AX322774.1	GI:18093754		
KEYWORDS				
SOURCE	unidentified.			
ORGANISM	unidentified.			
REFERENCE	1	unclassified.		
AUTHORS	Bunk, D., Reuner, B., Beck, J. and Henkel, T.			
TITLE	Novel target genes for diseases of the heart			
JOURNAL	Patent: WO 0192567-A 18 06-DEC-2001;			
FEATURES	Medigene AG (DE)			
Source	Location/Qualifiers			
	1..886			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
BASE COUNT	278 a	172 c	191 g	245 t
ORIGIN				
Query Match	100.0%;	Score 149;	DB 6;	Length 886;
Best Local Similarity	100.0%;	Pred. No. 8.8e-33;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GGGTTCCTCCCACTTCGGATGAGAGAAAGCCCAATTCGAGAGCGAAGAACTTCCA	60	
Db	316	GGGTTCCTCCCACTTCGGATGAGAGAAAGCCCAATTCGAGAGCGAAGAACTTCCA	375	
Qy	61	GGACCTGACAGTCAATCTATCGGAAATCCGAAATATTAAAGTAACTTAATATGTCCCC	120	
Db	376	GGACCTGACAGTCAATCTATCGGAAATCCGAAATATTAAAGTAACTTAATATGTCCCC	435	
Qy	121	AAAGCTGAACAGTAGTAGAGAAAGAAAG	149	
Db	436	AAAGCTGAACAGTAGTAGAGAAAGAAAG	464	
RESULT 5				
AF129505				
LOCUS	AF129505	886 bp	mRNA	linear
DEFINITION	Homo sapiens small muscular protein (SMPX)			
ACCESSION	AF129505			
VERSION	AF129505.1	GI:6625646		
KEYWORDS				
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
	Ekaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1	(bases 1 to 886)		
AUTHORS	Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.			
TITLE	Identification, mapping, and genomic structure of a novel			

FEATURES	source	location/Qualifiers
JOURNAL MEDLINE	X-chromosomal human gene (SMPX) encoding a small muscular protein	
PUBMED	Hum. Genet. 105 (5), 506-512 (1999)	
REFERENCE	2 (bases 1 to 886)	
AUTHORS	Patzak, D.	
TITLE	Direct Submission	
JOURNAL	Submitted (18-JAN-1999) Molecular Human Genetics, Institut for Human Genetics, Fleischmannstr. 42/44, D-17487 Greifswald, Germany	
gene	1..886	
5'UTR	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /map="Xq22.1-22.2" 1..>886 /gene="SMPX" 1..183 /gene="SMPX" 184..450 /gene="SMPX" /codon_start=1 /evidence=not_experimental /product="small muscular protein" /protein_id="AAFI9343.1" /db_xref="GI:6625647" /translation="MNMSKQPVSVNVRATQANINIPMGAFRGAGQPPRRKCTPEVEHGVPTSDKKKPIRGAKKLPGPYNVLSIQNIKSELKIVPAEQ"	
CDS	190 /gene="SMPX" /note="alternate position for initiation methionine" 451..>886 /gene="SMPX" 857..862 /gene="SMPX"	
misc_feature	279 a 172 c 191 g 244 t	
3'UTR		
polyA_signal		
BASE COUNT	279 a 172 c 191 g 244 t	
ORIGIN		
Query Match	100.0%; Score 149; DB 9; Length 886;	
Best Local Similarity	100.0%; Pred. No. 8 8e-33;	
Matches 149; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 GGTGTTCTCTCCCACTCGGATGAGGAGAAAGCCAAATTCAGAGCGAAGAACTTCCA 60	
DB	316 GGTGTTCTCTCCCACTCGGATGAGGAGAAAGCCAAATTCAGAGCGAAGAACTTCCA 375	
QY	61 GGACCTGCAGTCAATCTATCGGAATCCAGAAATTTAAATGAATTAATATATCTCCC 120	
DB	376 GGACCTGCAGTCAATCTATCGGAATCCAGAAATTTAAATGAATTAATATATCTCCC 435	
QY	121 AAAGCTGAACAGTAGTAGGAAGAAAAAG 149	
DB	436 AAAGCTGAACAGTAGTAGGAAGAAAAAG 464	
RESULT 6		
LOCUS	U73508 36503 bp DNA linear PRI 27-APR-1999	
DEFINITION	Homo sapiens cosmid clone U112E8 from Xp22.1-22.2, complete	
ACCESSION	U73508	
VERSION	U73508.1	
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 36503)	
TITLE	Toward a complete human genome sequence	
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)	
REFERENCE	99063792	
PUBMED	9847074	

REFERENCE 2 (bases 1 to 36503)
 AUTHORS Waterston,R.
 TITLE Unpublished (1999)
 JOURNAL 3 (bases 1 to 36503)
 REFERENCE 3 (bases 1 to 36503)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-1996)
 REFERENCE 4 (bases 1 to 36503)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (20-JAN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 36503)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (27-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT SUBMITTED BY: WUSGC
 Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 This clone was mapped by Grieff, M., Whyte, M. P., Thakker, R. V., and Mazzarella, R. Sequence analysis of 139 kb in Xp22.1 containing spermine synthase and the 5' region of PRX. Genomics 44:227-231 (1997).

SOURCE INFORMATION:
 This clone is from a chromosome X-specific cosmid library LL0XNCC01 'U'. The source of the chromosomes was a human/hamster hybrid, GM07297-F, from Robert Nussbaum at the University of Pennsylvania School of Medicine. Please contact the Lawrence Livermore National Laboratory at http://www-bio.llnl.gov/genome to obtain the clone.
 VECTOR: Lawriest16.
 Location/Qualifiers
 1..36503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="Xp22.1-22.2"
 /clone="U112E8"
 /clone_11b="LL0XNCC01-U"
 116..196
 /rpt_family="L2"
 1980..2046
 /rpt_family="L2"
 3690..4103
 /note="match to EST AA211443 (NID:G1810130) zn55B01.s1"
 3711..3929
 /note="match to EST AA214031 (NID:G1812669) zn58F02.s1"
 complement(3980..4103)
 /note="match to EST AA211521 (NID:G1810175) zn55B01.r1"
 4141..4364
 /rpt_family="L1"
 4366..4730
 repeat_region

/rpt_family="MaLR"
 4732..4953
 /rpt_family="L1"
 repeat_region
 5596..5711
 /rpt_family="MaLR"
 repeat_region
 6433..6793
 /rpt_family="L1"
 repeat_region
 7169..7254
 /rpt_family="L2"
 repeat_region
 7874..8264
 /rpt_family="L1"
 repeat_region
 8529..8555
 /rpt_family="L2"
 repeat_region
 10938..11032
 /rpt_family="MIR"
 repeat_region
 11095..11460
 /rpt_family="MaLR"
 repeat_region
 11640..11719
 /rpt_family="L1"
 repeat_region
 12097..12191
 /rpt_family="L1"
 repeat_region
 12453..12595
 /rpt_family="L1"
 repeat_region
 13060..13322
 /rpt_family="MER1_type"
 repeat_region
 13286..13385
 /rpt_family="MER1_type"
 repeat_region
 13401..13625
 /rpt_family="MER1_type"
 repeat_region
 14107..14274
 /rpt_family="MIR"
 repeat_region
 14638..14751
 /rpt_family="L1"
 repeat_region
 14804..15170
 /rpt_family="MaLR"
 repeat_region
 15277..15396
 /rpt_family="L2"
 repeat_region
 15512..15625
 /rpt_family="L2"
 repeat_region
 16941..17238
 /rpt_family="Alu"
 repeat_region
 17746..18251
 /rpt_family="MaLR"
 repeat_region
 19251..19321
 /rpt_family="MER1_type"
 repeat_region
 19583..19881
 /rpt_family="Alu"
 repeat_region
 20137..20152
 /rpt_family="L1"
 repeat_region
 20198..20256
 /rpt_family="MER1_type"
 repeat_region
 20456..20562
 /rpt_family="MIR"
 repeat_region
 20820..21010
 /rpt_family="MER1_type"
 repeat_region
 23777..23857
 /rpt_family="MIR"
 repeat_region
 24315..24471
 /rpt_family="MER1_type"
 repeat_region
 24480..24571
 /rpt_family="MIR"
 repeat_region
 24736..24842
 /rpt_family="L2"
 repeat_region
 26372..26608
 /rpt_family="MIR"
 repeat_region
 27011..27234
 /rpt_family="MER1_type"
 repeat_region
 28149..28286
 /rpt_family="MIR"
 repeat_region
 28660..29064
 /rpt_family="L1"
 repeat_region
 29526..29750
 /rpt_family="L1"
 repeat_region

```
repeat_region 30427. 30777 /rpt_family="MELR"
repeat_region 30958. 30968 /rpt_family="MERL_type"
repeat_region 34157. 34277 /rpt_family="MIR"
repeat_region 34281. 34328 /rpt_family="L2"
misc_feature 35221. 35386 /note="match to EST AA211443 (NID:g1810130) zn55b01.81"
misc_feature 35221. 35386 /note="match to EST AA211521 (NID:g1810175) zn55b01.r1"
misc_feature 36277. 36501 /note="match to EST AA389647 (NID:g2042633)"
repeat_region 36277. 36501 /rpt_family="Retroviral"
BASE COUNT 10773 a 7354 c 7510 g 10866 t
ORIGIN
```

Query Match	100.0%	Score 149,	DB 9,	Length 36503,
Best Local Similarity	100.0%	Prod. No.	8.7e-33,	
Matches 149, Conservative	0,	Mismatches	0,	Gaps 0,

QY	1	GGGTTTCTCCCACTCGGATGAGGAGAAAGCAATTCAGAGCGCAAGAACTTCCA	60
Db	35384	GGGTTTCTCCCACTCGGATGAGGAGAAAGCAATTCAGAGCGCAAGAACTTCCA	35325

```

07 61 GGACCTGAGTCAATCTATCGGAATCCAGAATTTAAAGTGAACCTAAATATGTCCC 120
    |||||
Db 35324 GGACCTGAGTCAATCTATCGGAATCCAGAATTTAAAGTGAACCTAAATATGTCCC 35265

```

```

QY      121  AAAGCTGAACAGTAGTAGGAAGAAAAAG  149
          |||||
Db      35264  AAAGCTGAACAGTAGTAGGAAGAAAAAG  35236

```

AL772370/c	LOCUS	DEFINITION	ACCESSION
AL772370	150319 bp	DNA 11near HTG 17-AUG-2002	
		Homo sapiens chromosome X clone RP1-184B10, *** SEQUENCING IN	
		PROGRESS ***, 2 unordered pieces.	
		AL772370	

```

VERSION      AL772370.5  GI:22416024
KEYWORDS     HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP
SOURCE       human.
ORGANISM     Homo sapiens

```

REFERENCE

Bates G
1964 C
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelcostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 150319)

TITLE Direct Submission
JOURNAL Submitted (15-AUG-2002) Wellcome Trust Sanger Institute, Hinxton
Cambridgehire, CB10 1SA, UK. E-mail enquiries:
humquerey@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Aug 21, 2002 this sequence version replaced gi:22204612.
..... Garcoo Garcoo

```

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA184B10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 99% of reads
Quality: 149793 bases at least Q40
Consensus quality: 149914 bases at least Q30
Consensus quality: 150006 bases at least Q20
Insert size: 150219; sum-of-contigs
Insert size: 152259; 2.5% error; agarose-fp
Quality coverage: 17.18x in Q20 bases; sum-of-contigs
Quality coverage: 17.58x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1      4859: contig of 4859 bp in length
*      *      4860 4959: gap of      100 bp
*      *      4960 150319: contig of 145360 bp in length.
FEATURES
      Location/Qualifiers
          1..150319
source

```

```

misc_feature
/note="assembly fragment:05115
fragment chain:1"
4960..150319
/note="assembly fragment:05270
fragment chain:1
clone end:SP6
vector_side:right"
47436 a 29741 c 29722 g 43320 t 100 othersg
BASE COUNT
ORIGIN
```

Query Match	100.0%	Score 149;	DB 2;	Length 150319;
Best Local Similarity	100.0%;	Prd. No. 8,6e-33;		
Matches 149;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	GGTGTTCTCCCACTCGGATAGAGAAAGAACCAATTCCAGACCGAAGAACTTCCA	60
Db	11854	GGTGTTCTCCCACTCGGATAGAGAAAGAACCAATTCCAGACCGAAGAACTTCCA	11799

Dy	61	GGACCGCAGTCATCTATCGAAATCCAGATATTAAAGTCACTAAAATATGTCCC	120
Dd	11794	GGACCTGCAGTCATCTATCGAAATCCAGATATTAAAGTGAACTAAAATATGTCCCC	11735

OY	121	AAAGCTGACAGTAGTAGGAAGAAAAAAG	149
Db	11734	AAAGCTGACAGTAGTAGGAAGAAAAAAG	11706

RESULT 8	AX332418/c	AX332418	587 bp	DNA	1linear	PAT 09-JAN-2002
LOCUS		Sequence				
DEFINITION		2927 from Patent				
ACCESSION		AX332418				
VERSION		AX332418.1				
KEYWORDS		GI:18123052				
SOURCE		human.				

CONVENTION	REFERENCE
homo sapiens	1
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	1
Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endreas, G., Horrihan, S., Soppet, D.R. and Weaver, Z.	1
Cancer gene determination and therapeutic screening using signature gene sets	1
Patent: WO 0194629-A 2927 13-DEC-2001;	1
Avalon Pharmaceuticals (US)	1
Location/Qualifiers	1

	source				
	1..587	/organism="Homo sapiens"			
BASE COUNT	177 a	/db_xref="taxon:9606"			
ORIGIN	122 c	103 g	183 t	2 others	
Query Match	75.7%	Score 112.8;	DB 6;	Length 587	

Query Match 75.7%; Score 112.8; DB 6; Length 587;

DY 68 CAGTCAATCTATTCGGAATCCAGAATAATTAAAGTGAACTTTAATATGTCCCCCAAAGCTG 127
|||||
DB 396 TTGTCAACTGTCTCGATCCAAAATGTTAAAGTGAACTGAAATTTGTCCCCCAAAGGTG 455

QY 128 AACAGTAGTAGGAAGAAAAA 148
|||
Db 456 AACAGTAGTCGAAAGGACACA 476

RESULT 10
AB3CA07A

LOCUS	AF364070	936 bp	linear	ROD 04-MAY-2001
DEFINITION	Mus musculus SMPX protein (Smpx)	mRNA, complete cds.		
ACCESSION	AF364070			
VERSION	AF364070.1	GI:13940507		

SOURCE ORGANISM

1 (bars 1 to 316)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus

REFERENCE	AUTHORS	TITLE
1 (cases 1, 2, 3, 5, 6)		
	Patzak, D., Zhuchenko, O., Lee, C.C. and Wehnert, M.	Identification, mapping, and genomic structure of a novel

JOURNAL
OF
HUMAN
GENETICS
20068878

MEASUREMENTS	20030675
PUBMED	10598820
REFERENCE	2 (bases 1 to 936)
AUTHORS	Patzak, D.
OTHERS	Direct Submission

FEATURES
Source 1 936
Location/Qualifiers
Fleischmannstr. 42-44, D-17487 Greifswald, Germany
Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Himnang.
Direct Submission
11100

gene

CDS	/gene=...smpx 199 . . 456 /gene="Smpx"
-----	--

```

/gene = smpx
/codon_start=1
/product="SMPX protein"

```

```

/protein_id="AAK50398.1"
/db_xref="GI:13940508"

```

```
/translation="MSKQPISNVRAIQININIMGAFRGAGQPPARKESIPEIEEG  
PTSEKKPIPGMKFPGPVNLSEIQNVKSELKFPVKGEQ"
```

```
polyA_signal 906. .911  
/gene="Smpx"  
/evidence=experimental]
```

BASE COUNT	283 a	187 c	203 g	263 t
ORIGIN				

Query Match 64.6%; Score 96.2; DB 10; Length 936

Best Local Similarity 80.1%; Pred. No. 1.7e-17;
Matches 113; Conservative 0; Mismatches 28; Indels 0; Gaps 0

8 CTCCACCTCGATGAGGAGGAGGCCAATTCAGGAGCGAAGAACTTCCAGGACCTG 67

Db 329 CTCTTACCACCTCAGAGGAAAAGAGCCCATTTCTGTGAATGAAGAATTTCCAGGACCTG 38

QY 68 CAGTCAATCTATCGGAATCCAGATATTAAAGTGAACCTAAAAATATGTCCCAAGCTG 12

Db 389 TTGTCACCTGTCTGAGATCCAAA

QY 128 AACAGTAGGAAGAAAA 148
|||||||

Db 449 AACAGTACTCGAAAGGACACA 469

RESULT 11
MATH047777

LOCUS	MMU245112
DEFINITION	MMU245112 943 bp mRNA linear ROD 12-APR-2001
DESCRIPTION	Mus musculus mRNA for stretch responsive muscle (X-chromosome)

ACCESSION AJ245772.1 GI:10178962
 VERSION AJ245772.1
 KEYWORDS Srmx gene; stretch responsive/muscle (X-chromosome).
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 943)
 AUTHORS Kemp,T.J., Sadusky,T.J., Simon,M., Brown,R., Eastwood,M., Sasason,D.A. and Coulton,G.R.
 TITLE Identification of a novel stretch-responsive skeletal muscle gene (Smrx)
 JOURNAL Genomics 72 (3), 260-271 (2001)
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2. (bases 1 to 943)
 AUTHORS Kemp,T.J.
 TITLE Direct Submission
 JOURNAL Submitted (12-AUG-1999) Kemp T.J., Molecular Pathology, Imperial College School Of Medicine, 6AV Building, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
 FEATURES
 source
 1..943
 /organism="Mus musculus"
 /strain="C57BL/10"
 /db_xref="taxon:10090"
 /tissue_type="skeletal muscle"
 1..943
 /gene="Smrx"
 1..214
 /gene="Smrx"
 215..472
 /gene="Smrx"
 /codon_start=1
 /product="stretch responsive muscle (X-chromosome)"
 /protein_id="CAC08493.1"
 /db_xref="GI:10178962"
 /translation="MSKQPISTNVAIININIPMGAFPGAGQPPRRKESTPETERGA PTSEKKPIPKMKFPVPVNUSEIQNVSLEKFPVKGSG"
 473..943
 /gene="Smrx"
 535..539
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 585..590
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 658..662
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 706..710
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 740..744
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 779..785
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 909..913
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 923..928
 /gene="Smrx"
 /function="mRNA deacabllising motif"
 943
 /gene="Smrx"
 BASE COUNT 278 a 192 c 210 g 262 t 1 others
 ORIGIN
 Query Match 64.6%; Score 96.2; DB 10; Length 943;
 Best Local Similarity 80.1%; Pred. No. 17e-17;
 Matches 113; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

07 8 CTCGCCCTCGATGAGAGAAAGCCAAATTCAGAGAGGAAACCTTCAGACCTG 67
 Db 345 CTCCTACCACTCGACAGAGAAAGAACCAATTCGAAAGAAATTTCCAGGACTG 404
 07 68 CAGTCATCTATCGAAATTCAGATATTAAAGTGAATTAATATATGTCGCCAAGCTG 127
 Db 405 TTGTCACTGTCTGAGATCCAAATGTAAAGTGAATCGAATTTGTCCTCAAGGCTG 464
 07 128 AACAGTAGTAGAAGAAAGAAA 148
 Db 465 AACAGTAGTAGAAGAGACAA 485
 RESULT 12
 AL732396 228031 bp DNA linear HTG 17-AUG-2002
 LOCUS Mus musculus chromosome X clone RP23-93M14, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 8 unordered pieces.
 AL732396 GI:22415936
 VERSION HTG; HTGS PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 228031)
 AUTHORS Heath,P.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SN, UK. E-mail enquiries: humquerry@sanger.ac.uk
 COMMENT On Aug 21, 2002 this sequence version replaced gi:22204493.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: Bm93M14
 ----- Summary Statistics
 Assembly program: XGAP4, version 4.5
 Chemistry: Dye-terminator; 99% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Consensus quality: 224691 bases at least Q40
 Consensus quality: 226148 bases at least Q30
 Consensus quality: 226956 bases at least Q20
 Insert size: 227331; sum-of-coverage
 Insert size: 194478; 8.5% error; agarose-fp
 Quality coverage: 6.42x in Q20 bases; sum-of-coverage
 coverage: 8.63x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 5544: contig of 5544 bp in length
 * 5545 5644: gap of 100 bp
 * 5645 11121: contig of 5477 bp in length
 * 11122 11221: gap of 100 bp
 * 11222 23110: contig of 11889 bp in length
 * 23111 23210: gap of 100 bp
 * 23211 42146: contig of 18936 bp in length
 * 42147 42246: gap of 100 bp
 * 42247 74925: contig of 32679 bp in length
 * 74926 75025: gap of 100 bp
 * 75026 153828: contig of 78803 bp in length
 * 153829 153928: gap of 100 bp
 * 153929 172635: contig of 18707 bp in length

FEATURES	Location/Qualifiers
source	1. .923 /organism="Xenopus laevis" /db_xref="taxon:83554"
gene	1. .923 /gene="Cg1"
CDS	435..710 /gene="Cg1" /codon_start=1 /product="Chisel" /protein_id="AAK71068.1" /db_xref="GI:14599748" /translation="MSKQPSNINISINAMINIPMGAFRPGAGPPRRKEFSTEEBQHV PRPESESEKKIPGAVNKIPGAPFALISIQNVKSLKVPFAEEQ"
BASE COUNT	312 a 184 c 187 g 240 t
ORIGIN	
Query Match	49.1%; Score 73.2; DB 5; Length 923; Best Local Similarity 76.3%; Pred. No. 7.9e-11; Matches 90; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
QY	21 TGAAGGAAGAACCAATTCGAGGCGAAGAACTTCAGAGCCTCACTCAATCATC 80
Db	593 TGAAGGAAGAACTTATTCGAGGCGAGTAAACTTCAGAGGCGCGCTCAACTCTC 652
QY	81 GGAATCCAGATATTAAAGTGAACCTAAATATCTCCCAAGCTAACAAGTAGT 138
Db	653 AGAATCCAGATGTAAAGTGTACTTAATTGTACCAAAAGCAGAGACATATG 710
RESULT 15	
LOCUS	ATL73G19 42446 bp DNA linear PLN 27-MAY-1999
DEFINITION	Arabidopsis thaliana DNA chromosome 4, cosmid clone L73G19 (ESSA project).
ACCESSION	AL050400
VERSION	AL050400.1
KEYWORDS	GI:4914454
SOURCE	Arabidopsis thaliana.
ORGANISM	Arabidopsis thaliana. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; 1 (bases 1 to 42446) Beyan,M., Pohl,T., Weizenegger,T., Bancroft,I., Mewes,H.W., Meyer,K.F.X. and Scheller,C. Unpublished 2 (bases 1 to 42446) EU Arabidopsis sequencing project. Direct Submission Submitted (27-MAR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schuelli@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Beyan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.beyan@bbsrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ . Location/Qualifiers 1. .42446
FEATURES	Location/Qualifiers
source	1. .42446 /organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" 1. .3310 /note="overlap to BAC M7J2; please refer to EMBL acc:AL022197 for analysis and annotation" 4971..6937 /gene="L73G19.10" join(4971..4989,5232..5572,5666..5761,5846..5923,6011..6187,6602..6670,6755..6937) /gene="L73G19.10"
misc_feature	
gene	
CDS	

```

/gene="L73G19.30"
/number=3
complement(12638.12835)
/gene="L73G19.30"
/number=4
complement(12836.13021)
/gene="L73G19.30"
/number=4
complement(13022.13318)
/gene="L73G19.30"
/number=5
complement(13319.13369)
/gene="L73G19.30"

exon
intron

intron

Query Match
Best Local Similarity 24.6%; Score 36.6; DB 8; Length 42446;
Matches 72; Conservative 55.0%; Pred.No. 3.1;
Matches 59; Indels 0; Gaps 0.

Db 40722 TTGGATTAAGCAGACGTAATAATTGATGTACCCAAAGTAAAGCTTAGGATCTGCACAACTA 40781
QY 16 TCGGATGAGAGAGACGCAATTCACAGAGCGAGAAACCTCCAGAGCCTGCAGTCAAT 75
Db 40722 TTGGATTAAGCAGACGTAATAATTGATGTACCCAAAGTAAAGCTTAGGATCTGCACAACTA 40781
QY 76 CTATCGGAATCCAGAACTTAATAAGTAGACTTAATAATATGTCCCAAGCTGAACAGTAG 135
Db 40782 ATGAGTAAATCTCAATATTTAACTCCGTTTAAGACACGGCCACTAATCAATAATGATTAA 40841
QY 136 TAGGAGAGAAA 146
Db 40842 AACCAATCAAGA 40852

Search completed: April 15, 2003, 20:59:14
Job time : 446.585 secs

```

COMMENT

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com).
 Seq primer: M13 Reverse.

FEATURES

Source

Location/Qualifiers

1..541

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJO-ahk-c-03-0-UI"

/clone_lib="UI-E-EJO"

/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"

/dev_stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-EJO is a subcloned cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTGG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCCA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT

188 a 82 c 100 g 171 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 422; DB 14; Length 541;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 8 GA 7

RESULT 2
 BM717052

LOCUS
 DEFINITION UI-E-EJO-ahk-c-03-0-UI. r2 UI-E-EJO Homo sapiens CDNA clone

ACCESSION
 VERSION BM717052

KEYWORDS
 SOURCE EST.

ORGANISM
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 541)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 GENOME RES. 6 (9), 791-806 (1996)
 MEDLINE 97044477

541 bp mRNA linear EST 28-FEB-2002

QY 1 GATTGATGGAAGAAATTAAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 60

DB 97 GATTGATGGAAGAAATTAAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 156

QY 61 TATTGTATGATGATGTGAACCTCCTGAATGCGTGAAGCTTACAGAAATGGCTGTT 120

DB 157 TATTGTATGATGATGTGAACCTCCTGAATGCGTGAAGCTTACAGAAATGGCTGTT 216

QY 121 TGTACATTTAATCTCTCTCTAGTGGCTGTATTTCTTACTTAACTTCAATTTTGG 180

DB 217 TGTACATTTAATCTCTCTCTAGTGGCTGTATTTCTTACTTCAATTTTGG 276

QY 181 CACCTCAGAGAAATTAAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 240

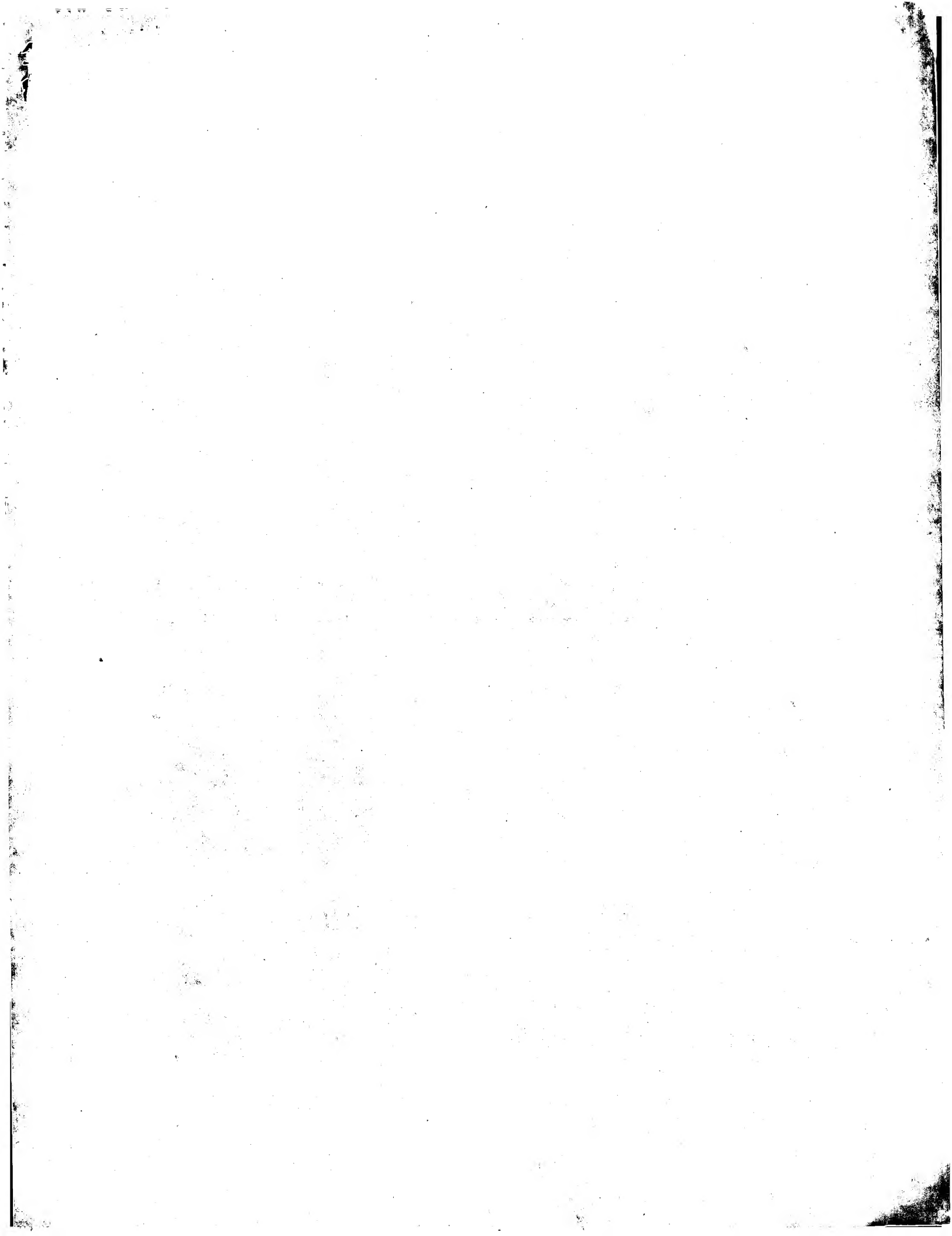
DB 277 CACCTCAGAGAAATTAAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 336

QY 241 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

DB 337 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 396

QY 301 CAACTACTTCTTACAGATTTTGTCAATTTTGAATGGCTTTAGTTTCACTTT 360

DB 397 CAACTACTTCTTACAGATTTTGTCAATTTTGAATGGCTTTAGTTTCACTTT 456



Qy 361 TAAATTATGCTAACTTTGTATGAGTTCATTAATATTTGACATAATGTAATG 420
 Db 457 TAAATTATGCTAACTTTGTATGAGTTCATTAATATTTGACATAATGTAATG 516
 Qy 421 GA 422
 Db 517 GA 518

518-97-4001

RESULT 3

BM674432/c

LOCUS 544 bp mRNA linear EST 27-FEB-2002

DEFINITION UI-E-EJO-ahk-c-03-0-UI.s1 UI-E-EJO Homo sapiens cDNA clone

ACCESSION BM674432

VERSION BM674432.1 GI:18984330

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Bonaldo, M.P., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.regen.com).

Seq primer: M13 Forward

POLY-A-Tes.

FEATURES

Source

Location/Qualifiers

1..544

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJO-ahk-c-03-0-UI"

/clone_1db="UI-E-EJO"

/tissue-type="fetal eyes, lens, eye anterior segment,

optic nerve, retina, Retina Foveal and Macular, RPE and

Choroid"

/dev stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site 1: EcoR I; Site 2: Not I;

UI-E-EJO is a subtracted cDNA library constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I adaptor, digested

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pT73-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tags for this library are: fetal eyes, AGATCAAGA

; lens, CGATTACCGA; eye anterior segment, AATGCCGAT;

Macular, GTCC; RPE and Choroid, ACCGA. This library was

created for the program, Gene Discovery in the Visual

System, supported by National Eye Institute (NEI).

Tag LIB=UI-E-EJO

Tag TISSUE=Foveal and Macular Retina

Tag SEQ=CTCC

BASE COUNT 171 a 103 c 82 g 188 t
 ORIGIN
 Query Match 100.0%; Score 422; DB 14; Length 544;
 Best Local Similarity 100.0%; Pred. No. 3.6e-83;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATTGATGTGAAGAAATTAAGAGCAGAGATGATGATTAATTAATTAATTA 60
 Db 439 GATTGATGTGAAGAAATTAAGAGCAGAGATGATGATTAATTAATTAATTA 380
 Qy 61 TATTGATGTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 120
 Db 379 TATTGATGTGAAGATGATGATGATGATGATGATGATGATGATGATGATG 320
 Qy 121 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
 Db 319 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 260
 Qy 181 CACTCAGAGACAAATTAGCCCATTAATTAACACCTGAGAGGAGGAGGAGG 240
 Db 259 CACTCAGAGACAAATTAGCCCATTAATTAACACCTGAGAGGAGGAGGAGG 200
 Qy 241 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATG 300
 Db 199 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATG 140
 Qy 301 CAAGTACTTCTCAGATGATGATGATGATGATGATGATGATGATGATGATG 360
 Db 139 CAAGTACTTCTCAGATGATGATGATGATGATGATGATGATGATGATGATG 80
 Qy 361 TAAATTATGCTAACTTTGTATGAGTTCATTAATATTTGACATAATGTAATG 420
 Db 79 TAAATTATGCTAACTTTGTATGAGTTCATTAATATTTGACATAATGTAATG 20
 Qy 421 GA 422
 Db 19 GA 18

RESULT 4

BE856875/c

LOCUS 683 bp mRNA linear EST 29-SEP-2000

DEFINITION IMAGE:3300004.3 Similar to TR:Q9DHP9 Q9DHP9 SMALL MUSCULAR

PROTEIN. mRNA sequence.

ACCESSION BE856875

VERSION BE856875.1 GI:10370341

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cga@nci.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40UP from G1bco

High quality sequence stop: 444.

Location/Qualifiers

1..683

/organism="Homo sapiens"

/db_xref="taxon:9606"

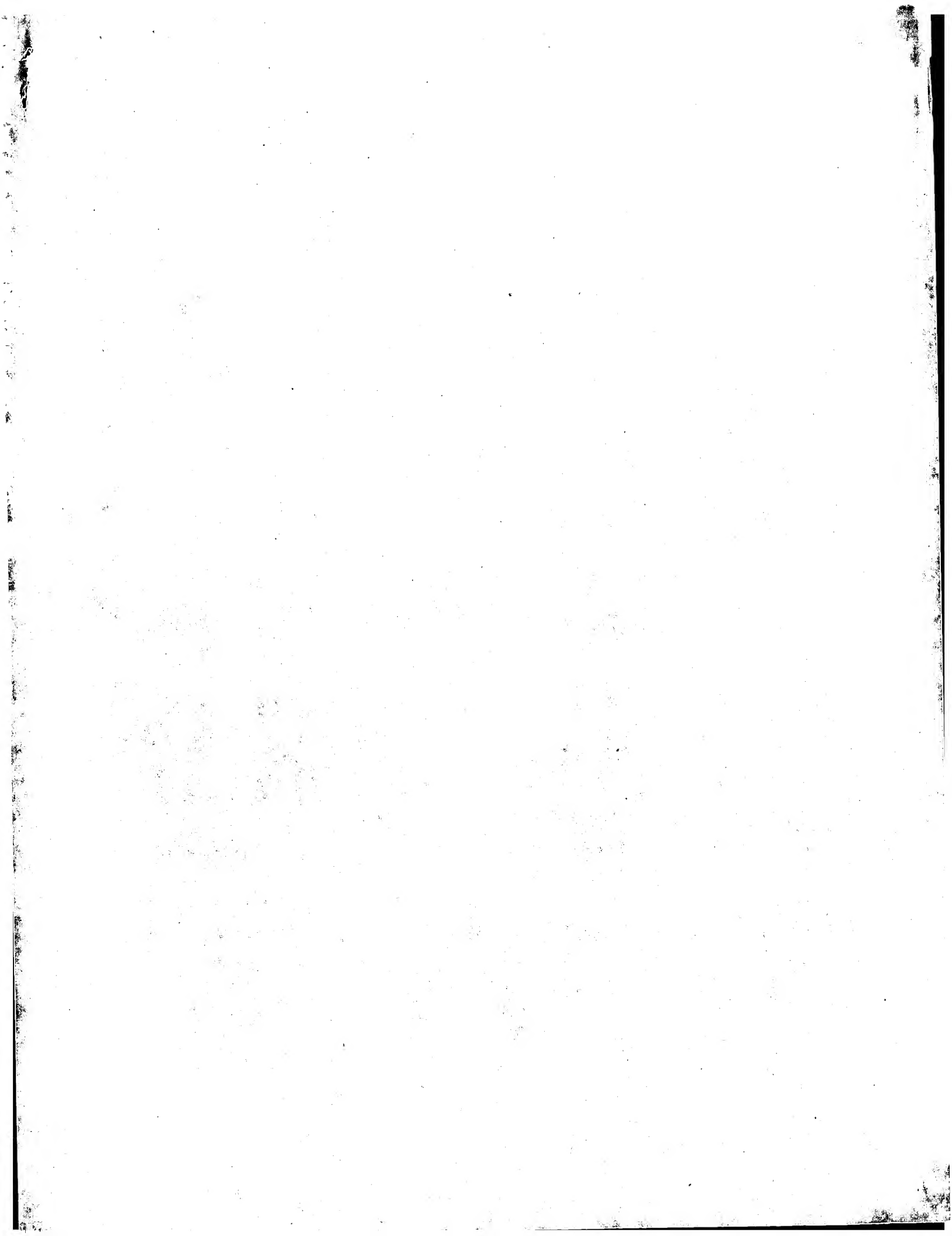
/clone="IMAGE:3300004"

/clone_1db="Soares_NSF_P8_9W_OT_PA_P_S1"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;



RESULT 7
AA211443/c
LOCUS
DEFINITION AA211443 587 bp mRNA linear EST 31-JAN-1997
ACCESSION ZN5501.81 Striatogene muscle 937209 Homo sapiens cDNA clone
IMAGE:562057 3', mRNA sequence.
VERSION AA211443
KEYWORDS AA211443.1 GI:1810130
EST.

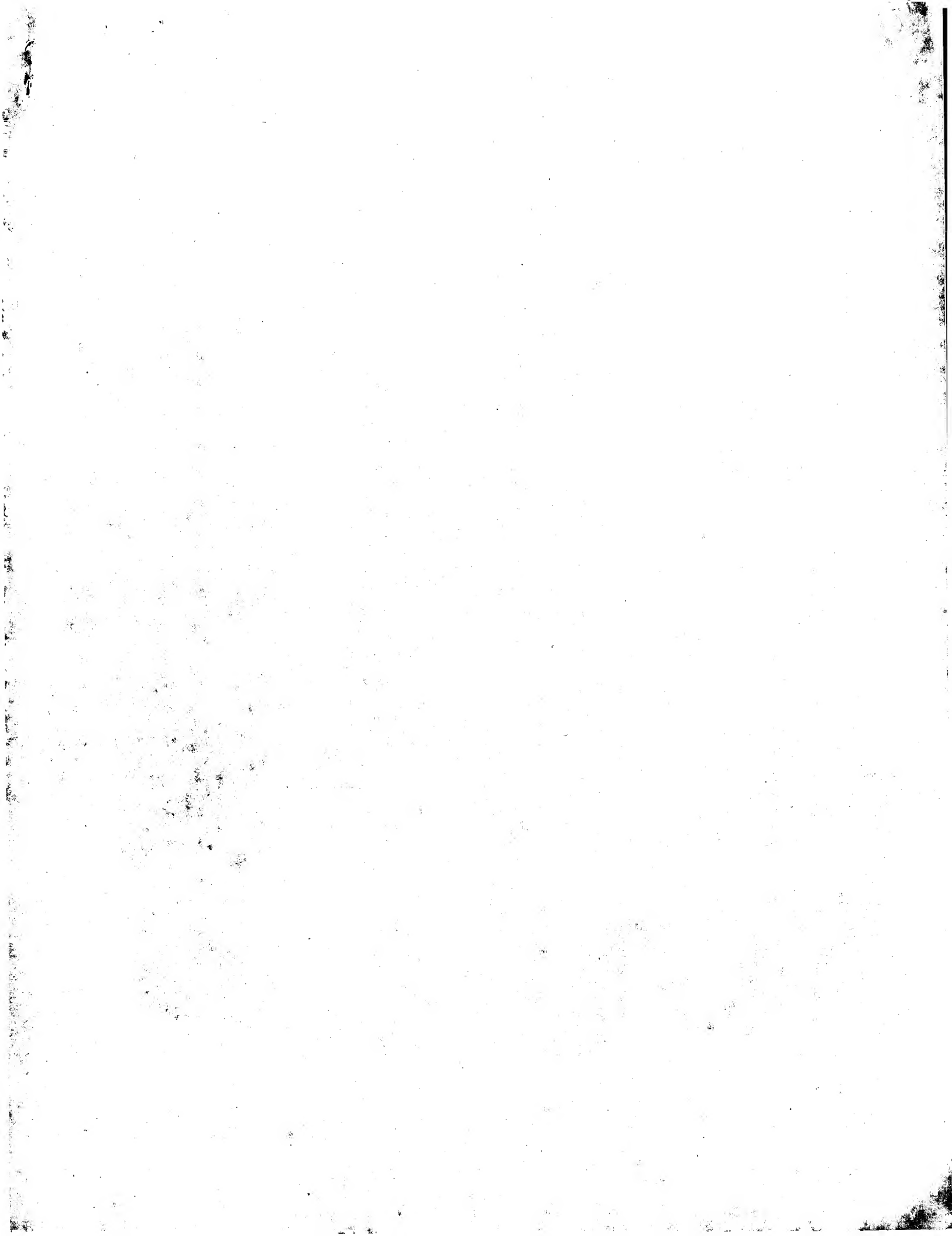
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 587)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoso, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m13 fwd. EF from Amersham
High quality sequence stop: 265.

FEATURES
source
1..587
/organism="Homo sapiens"
/db_xref="GDB:459347"
/db_xref="taxon:9606"
/clone="IMAGE:562057"
/clone_lib="Striatogene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Skeletal muscle from patient with malignant hyperthermia. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGCAGG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 177 a 122 c 103 g 183 t 2 others
ORIGIN

Query Match 97.6%; Score 412; DB 9; Length 587;
Best Local Similarity 100.0%; Pred. No. 5.6e-81;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGGAGAGAGATTCATCAATAGCTCACTAAATTTTATA 60
DB 412 GATTGATGTGAAGAAATTAAGAGGAGAGAGATTCATCAATAGCTCACTAAATTTTATA 353
QY 61 TATTTGTATGATGATGTAAGACCTTCCTGAATGCTGAGACTCTGACGAATAGCGCTGT 120
DB 352 TATTTGTATGATGATGTAAGACCTTCCTGAATGCTGAGACTCTGACGAATAGCGCTGT 293
QY 121 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCATTTTGG 180
DB 292 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCATTTTGG 233
QY 181 CACCTCAGAGAACAATTAAGCCATTAATTCACACCTGAGGGTGTGTTTGAGAGAG 240
DB 232 CACCTCAGAGAACAATTAAGCCATTAATTCACACCTGAGGGTGTGTTTGAGAGAG 173
QY 241 GATATGATTTTATGAGAGATATATGCAATGTCCTTAACATTTTATGATAAAGTTTCC 300
DB 172 GATATGATTTTATGAGAGATATATGCAATGTCCTTAACATTTTATGATAAAGTTTCC 113
QY 301 CAAGCTACTTCTACAGATTTTGTCAATATTTGGAATGCGTTTATGTTCTTACCTTT 360
DB 112 CAAGCTACTTCTACAGATTTTGTCAATATTTGGAATGCGTTTATGTTCTTACCTTT 53
QY 361 TAAATATGCTACCTAACTTTGTATAGTTCAATTAATTAATTTGACCTAATG 412



Wed Apr 16 05:43:02 2003

us-09-6

Db 52 TAAATTATGCTAACTTGTAGTCAATAAATATTGACTAATG 1

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:52:31 ; Search time 875.253 Seconds
(without alignment)
7808.593 Million cell updates/sec

Title: US-09-647-019-10

Perfect score: 422
Sequence: 1 gactgacgtgaaagaataaa.....tgactcaatgcaaatgca 422

Scoring table:

IDENTITY NUC
Gapop 10-0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_ests:
2: em_ests:
3: em_ests:
4: em_ests:
5: em_ests:
6: em_ests:
7: em_ests:
8: em_ests:
9: gb_ests:
10: gb_ests:
11: gb_ests:
12: gb_ests:
13: gb_ests:
14: gb_ests:
15: em_ests:
16: em_ests:
17: gb_ests:
18: em_ests:
19: em_ests:
20: em_ests:
21: em_ests:
22: em_ests:
23: em_ests:
24: em_ests:
25: em_ests:
26: em_ests:
27: em_ests:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422	100.0	455	9	A1355676
2	422	100.0	541	14	BM717052
3	422	100.0	544	14	BM674432
4	417	98.8	683	12	BM674432
5	413	97.9	495	9	BM674432
6	412	97.6	452	9	A1090520
					A1190689

7	412	97.6	587	9	AA211443
8	411	97.4	439	9	A1333485
9	411	97.4	494	10	AM170011
10	382	90.5	683	12	BF111459
11	316.2	74.9	368	14	BF111459
12	300.8	71.3	321	9	A1824667
13	298.4	70.7	313	14	F269162
14	291.4	69.1	302	14	F311142
15	290.6	68.9	719	14	BM697544
16	260	61.6	260	9	A1355906
17	254.6	60.3	349	9	AU277456
18	229	54.3	240	14	F16837
19	219.6	52.0	657	14	BQ176246
20	219	51.9	499	12	BG224307
21	215.6	51.1	557	13	BM123288
22	215.6	51.1	578	14	BQ554132
23	215.6	51.1	934	11	AK010172
24	214.6	50.9	890	9	AV088480
25	210.4	49.9	461	9	AA901082
26	210.4	49.9	498	9	A1574754
27	210.2	49.8	498	10	AM918887
28	209.6	49.7	393	9	AV006796
29	209.6	49.7	855	12	BF672902
30	208.4	49.4	534	9	A1577882
31	207.2	49.1	399	9	A1579785
32	207	49.1	521	9	AV005538
33	205.4	48.7	436	12	BF555737
34	203.2	48.2	756	12	BF791178
35	202.2	47.9	393	9	A1713293
36	202.8	47.6	934	11	AK003105
37	199.6	47.3	220	9	AA214031
38	193	45.7	330	9	A1355903
39	193	45.7	330	9	A1355904
40	191	45.3	191	9	A1381720
41	182	43.1	582	12	BG224246
42	181.6	43.0	371	10	BE111545
43	178	42.2	400	9	AV011785
44	177	41.9	472	9	A1232234
45	176.8	41.9	371	10	BE111689

ALIGNMENTS

RESULT 1
A1355676
LOCUS
DEFINITION
IMAGE:1952006 3', mRNA sequence.
ACCESSION
A1355676
VERSION
A1355676.1 GI:4095829
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 455)
AUTHORS
NCI-CCAP
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strauberg, Ph.D.
Email: cgap@remail.nih.gov
This clone is available royalty-free through LNL, contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 438.
Location/Qualifiers
1. 455
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1952006"

```

/clone_lib="Soares_fetal_lung_NBH19W"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: lung; Vector: pT73D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-GTGTACCATCTGAGAGTGGAGCGGCCGCAATTTTCTTTTCTTTT-3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Ronaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NBH19W."

BASE COUNT      155 a      88 c      65 g      147 t
ORIGIN

Query Match      100.0%; Score 422; DB 9; Length 455;
Best Local Similarity 100.0%; Pred. No. 3.7e-83;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTGATGTGAAGAATAAAGAGCAGAAAGATGATTCATATGCTCACTAAATTTTATA 60
DB 428 GATTGATGTGAAGAATAAAGAGCAGAAAGATGATTCATATGCTCACTAAATTTTATA 369
OY 61 TATTTGTATGATGATTTGGAACCTCTGAATGCTTAGACAGAAATGGCTGTT 120
DB 368 TATTTGTATGATGATTTGGAACCTCTGAATGCTTAGACAGAAATGGCTGTT 309
OY 121 TGTACATTTATATCTCTCTCTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGG 180
DB 308 TGTACATTTATATCTCTCTCTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGG 249
OY 181 CACCTTCACGAACAATTTAGCCCTTAATTTCAACCTCGAGAGGCTGTTTGGAGAG 240
DB 248 CACCTTCACGAACAATTTAGCCCTTAATTTCAACCTCGAGAGGCTGTTTGGAGAG 189
OY 241 GATATGATTTTATGAGATGATGATGCAATGCTTACGATTTTGAATGAAAATTTCC 300
DB 188 GATATGATTTTATGAGATGATGCAATGCTTACGATTTTGAATGAAAATTTCC 129
OY 301 CAAGCTACTTCTCTCAAGATTTTGTCAATTTTGAATGCGTTTACTTCACTTC 360
DB 128 CAAGCTACTTCTCTCAAGATTTTGTCAATTTTGAATGCGTTTACTTCACTTC 69
OY 361 TAAATATGCTCAATTAAGTGTATGATGTTCAATTAATTTGAATGTTAAATGT 420
DB 68 TAAATATGCTCAATTAAGTGTATGATGTTCAATTAATTTGAATGTTAAATGT 9
OY 421 GA 422
DB 8 GA 7

RESULT 2
BM717052 541 bp mRNA linear EST 28-FEB-2002
LOCUS UI-E-EJ0-ahk-c-03-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
DEFINITION UI-E-EJ0-ahk-c-03-0-UI 5', mRNA sequence.
ACCESSION BM717052
VERSION BM717052.1 GI:19030310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Chordata; Craniata; Vertebrata; Euteleostomi;
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 541)
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

```

COMMENT

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hegeman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

1..541

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ahk-c-03-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RRB and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Ronaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)₁₈ tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT;
; optic nerve, CCATTAGAGT; retina, ACCCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

```

BASE COUNT      188 a      82 c      100 g      171 t
ORIGIN

Query Match      100.0%; Score 422; DB 14; Length 541;
Best Local Similarity 100.0%; Pred. No. 3.6e-83;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATTGATGTGAAGAATAAAGAGCAGAAAGATGATTCATATGCTCACTAAATTTTATA 60
DB 97 GATTGATGTGAAGAATAAAGAGCAGAAAGATGATTCATATGCTCACTAAATTTTATA 156
OY 61 TATTTGTATGATGATTTGGAACCTCTGAATGCTTAGACAGAAATGGCTGTT 120
DB 157 TATTTGTATGATGATTTGGAACCTCTGAATGCTTAGACAGAAATGGCTGTT 216
OY 121 TGTACATTTATATCTCTCTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGG 180
DB 217 TGTACATTTATATCTCTCTCTAGTGGCTGATTTCTTACTTATCTTCAATTTTGG 276
OY 181 CACCTTCACGAACAATTTAGCCCTTAATTTCAACCTCGAGAGGCTGTTTGGAGAG 240
DB 277 CACCTTCACGAACAATTTAGCCCTTAATTTCAACCTCGAGAGGCTGTTTGGAGAG 336
OY 241 GATATGATTTTATGAGATGATGCAATGCTTACAGATTTTGAATGTTTCC 300
DB 337 GATATGATTTTATGAGATGATGCAATGCTTACAGATTTTGAATGTTTCC 396
OY 301 CAAGCTACTTCTCTCAAGATTTTGTCAATTTTGAATGCGTTTACTTCACTTT 360
DB 397 CAAGCTACTTCTCTCAAGATTTTGTCAATTTTGAATGCGTTTACTTCACTTT 456

```

QY 361 TAAATTATGCTAAACCTTGATGATGCTCAATAAATTTGACTAAATGTAATGT 420
 DB 457 TAAATTATGCTAAACCTTGATGATGCTCAATAAATTTGACTAAATGTAATGT 516
 QY 421 GA 422
 DB 517 GA 518

518-47441

RESULT 3
 BM674432/c
 LOCUS 544 bp mRNA linear EST 27-FEB-2002
 DEFINITION UI-E-E30-ahk-c-03-0-UI-81 UI-E-E30 Homo sapiens cDNA clone
 ACCESSION BM674432
 VERSION BM674432.1 GI:18984330
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 544)
 REFERENCE
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL
 MEDLINE
 COMMENT
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward
 POLYA-yes.

FEATURES

source

Location/Qualifiers
 1..544
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="UI-E-E30-ahk-c-03-0-UI"
 /clone_1lb="UI-E-E30"
 /tissue_type="fetal eyes, lens, eye anterior segment,
 optic nerve, retina, Retina foveal and Macular, RPE and
 Choroid"
 /dev_stage="fetal and adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-E30 is a subcloned cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into pT73-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for this library are: fetal eyes, AGAATCAGA
 / lens, CGATTAGCGA; eye anterior segment, AATCCGAT;
 optic nerve, CCAATAGTG; retina, CCGCG; Retina foveal and
 Macular, GTCC; RPE and Choroid, ACCCA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI).
 TAG_LIB=UI-E-E30
 TAG_TISSUE=Foveal and Macular Retina
 TAG_SEQ=GTCC"

BASE COUNT 171 a 103 c 82 g 188 t
 ORIGIN
 Query Match 100.0%; Score 422; DB 14; Length 544;
 Best Local Similarity 100.0%; Pred. No. 3.6e-83;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATTTGAAGAAATPAAAGGCGAAGATGATGATCAATAGCTCAATAATTTTATA 60
 DB 439 GATTGATTTGAAGAAATPAAAGGCGAAGATGATGATCAATAGCTCAATAATTTTATA 380
 QY 61 TATTGATGATGATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 379 TATTGATGATGATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 320
 QY 121 TGATGATTTATATCT 180
 DB 319 TGATGATTTATATCT 260
 QY 181 CACCTCAGAAACAAATTTAGCCCTTAATTCACACCTGAGGAGGTGTTGAGAGG 240
 DB 259 CACCTCAGAAACAAATTTAGCCCTTAATTCACACCTGAGGAGGTGTTGAGAGG 200
 QY 241 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 199 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 140
 QY 301 CAACTACTCTCTACAGATTTTGTCTAATTTTGGATGATGATGATGATGATGATGATGAT 360
 DB 139 CAACTACTCTCTACAGATTTTGTCTAATTTTGGATGATGATGATGATGATGATGATGAT 80
 QY 361 TAAATTATGCTAAACCTTGATGATGCTCAATAAATTTGACTAAATGTAATGT 420
 DB 79 TAAATTATGCTAAACCTTGATGATGCTCAATAAATTTGACTAAATGTAATGT 20
 QY 421 GA 422
 DB 19 GA 18

RESULT 4
 BE856875/c
 LOCUS 683 bp mRNA linear EST 29-SEP-2000
 DEFINITION 7F70G03.X1 Soares NSP_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:3300004 3' similar to TR:Q9UHP9 Q9UHP9 SMALL MUSCULAR
 PROTEIN /; mRNA sequence.
 BE856875
 ACCESSION BE856875.1 GI:10370341
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 683)
 REFERENCE
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 COMMENT
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgsapb-remail.nih.gov
 This clone is available royalty-free through LNL, contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40up from GIBCO
 High quality sequence stop: 444.

FEATURES

source

Location/Qualifiers
 1..683
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3300004"
 /clone_1lb="Soares NSP_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;

```

1.495
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1692097"
/clone_lib="Soares_fetal_heart_NBDH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notice="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGGAAGTGGAGCGCCGCGCATCTTTTCTTTTCTTTT 3']"
double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into

```

adapters (Pharmacia), digested with Not I and *Sal* I and cloned into the Not I and Eco RI sites of a modified pTV73 vector (Pharmacia). Library went through one round of normalization to a $Cot = 5$. Library constructed by M. Fatima Bonardo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH149W."

Query Match	97.9%;	Score 413;	DB 9;	Length 495;
Best Local Similarity	100.0%;	Pred. NO. 3.5e-81;		
Matches 413;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0
QY 1	GATTGATGTGAAGAAATTAAGAGGACGAAGATGTGATTCATTAATAGCTCACTAAAATTTTATA	60		
Db 413	GATTGATGTGAAGAAATTAAGAGGACGAAGATGTGATTCATTAATAGCTCACTAAAATTTTATA	354		
QY 61	TATTTGTATGATGATTTGTGAACCTCTCTGATATGCTTGAGACTCTTGACGAAGATGGCCGTGT	120		
Db 353	TATTTGTATGATGATTTGTGAACCTCTCTGATATGCTTGAGACTCTTGACGAAGATGGCCGTGT	294		
QY 121	TGTACATTTATATCTCTCTCTTAAGTTGGGCGTATTTCTTACTTTATCTTCATTTTGG	180		
Db 293	TGTACATTTATATCTCTCTCTTAAGTTGGGCGTATTTCTTACTTTATCTTCATTTTGG	234		
QY 181	CACCTCACAGAAACAATTAGCCCATTAATTCACAACCTGAGGGGTGTGGTTTGGAGAGG	240		
Db 233	CACCTCACAGAAACAATTAGCCCATTAATTCACAACCTGAGGGGTGTGGTTTGGAGAGG	174		
QY 241	GATATGATTTTATGAGAGATGATATGGCAATGTGCTTACAGATTTTGATGAAAAAGTTTCC	300		
Db 173	GATATGATTTTATGAGAGATGATATGGCAATGTGCTTACAGATTTTGATGAAAAAGTTTCC	114		
QY 301	CAACCTCTCTCTCAAGTATTTTGTGACATATTTGGATATTTGGATATGGTTTATGTTTCACTTT	360		

DB	113	CAGGCTACTTCCACAGAATTTTGTCACAAATTGGCAAAAATTAAGGAATGCATTATACTCCTCAT	
OY	361	TAAATTATGTCACTAACAATTGTATGAGTTCAAATTAATATTTGACTAAATGT	413
Ddb	53	TAAATTATGTCACTAACAATTGTATGAGTTCAAATTAATATTTGACTAAATGT	1
RESULT 6			
AII90689/c			
LOCUS	AII90689	452 bp	mRNA linear EST 28-OCT-1998
DEFINITION	qds0d0t.xl Soares_fetal_heart NBHHLIHW Homo sapiens cDNA clone.		
IMAGE:	J732897 3' , mRNA sequence.		
ACCESSION	AII90689		
VERSION	AII90689.1	GI:3741898	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria;	Primates; Catarrhini; Hominoidea; Homo.		
(bases 1 to 452)			
NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap .		
National Cancer Institute,	Cancer Genome Anatomy Project (CGAP),		
Tumor Gene Index			

Db 52 TAAATTATGTCACCTAATTGATGAGTTCATAATAATATTGACTAATG 1

RESULT 8
 AI333485 439 bp mRNA linear EST 13-FEB-1999
 LOCUS CP88f03.x1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
 DEFINITION IMAGE:1930109 3', mRNA sequence.
 ACCESSION AI333485
 VERSION AI333485.1 GI:4070044
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 439)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 851 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 423.

FEATURES
 source
 1..439
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1930109"
 /clone_lib="Soares_fetal_lung_NbHL19W"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: lung; Vector: pRT73D (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCAATTTTCTTTT-3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pRT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by Bento
 Soares and M. Fatima Bonaldo. This library was constructed
 from the same fetus as the fetal heart library, Soares
 fetal heart NbHL19W."

BASE COUNT 152 a 86 c 65 g 135 t 1 others
 ORIGIN

Query Match 97.4%; Score 411; DB 9; Length 439;
 Best Local Similarity 99.8%; Pred. No. 9.7e-81;
 Matches 411; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGAGAGATGATGATCTGCTCAATTTTATA 60
 Db 412 GATTGATGTGAAGAAATTAAGAGAGAGATGATGATCTGCTCAATTTTATA 353

QY 61 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 Db 352 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

QY 121 TGTACATTTATCT 180
 Db 292 TGTACATTTATCT 233

QY 181 CACCTCAGAGAAATTAAGCCATTAATCAACCTGAGGGGTGTGTTTGAGAGG 240
 Db 232 CACCTCAGAGAAATTAAGCCATTAATCAACCTGAGGGGTGTGTTTGAGAGG 173

QY 241 GATATGATTTTATGAGAGATATGAGCAATGTGCTTAACGATTTTATGATAAGTTTC 300
 Db 172 GATATGATTTTATGAGAGATATGAGCAATGTGCTTAACGATTTTATGATAAGTTTC 113

QY 301 CAGGCTACTCTCTACAGATTTTGGTCAATTTTGAATGGCTTTAGTCTTACCTT 360
 Db 112 CAGGCTACTCTCTACAGATTTTGGTCAATTTTGAATGGCTTTAGTCTTACCTT 53

QY 361 TAAATTATGTCACCTAATTGATGAGTTCATAATAATATTGACTAATG 412
 Db 52 TAAATTATGTCACCTAATTGATGAGTTCATAATAATATTGACTAATG 1

RESULT 9
 AM170011/c 494 bp mRNA linear EST 12-NOV-1999
 LOCUS Xj35g08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
 DEFINITION IMAGE:2659262 3', mRNA sequence.
 ACCESSION AM170011
 VERSION AM170011.1 GI:6401536
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 494)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 449.

FEATURES
 source
 1..494
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2659262"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pRT73D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NbHL19W, testis NHT, and B-cell
 NCI CGAP GCBI) were mixed, and 88 circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 1.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 165 a 93 c 76 g 159 t 1 others
 ORIGIN

Query Match 97.4%; Score 411; DB 10; Length 494;
 Best Local Similarity 99.8%; Pred. No. 9.6e-81;
 Matches 411; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGAGAGATGATGATCTGCTCAATTTTATA 60
 Db 412 GATTGATGTGAAGAAATTAAGAGAGAGATGATGATCTGCTCAATTTTATA 353

QY 61 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 Db 352 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

QY 121 TGTACATTTATCT 180
 Db 292 TGTACATTTATCT 233

QY 181 CACCTCAGAGAAATTAAGCCATTAATCAACCTGAGGGGTGTGTTTGAGAGG 240
 Db 232 CACCTCAGAGAAATTAAGCCATTAATCAACCTGAGGGGTGTGTTTGAGAGG 173

Db 232 CACCTCAGAGAACTTACCCCTAAATTCACACCTCGAGGCTGTGTTTGAAGAG 173
 Qy 241 GATATGATTTTATGAGAAATGATATGCGCAATGCTCTTAACGATTTTGTATGAAGTTCC 300
 Db 172 GATATGATTTTATGAGAAATGATATGCGCAATGCTCTTAACGATTTTGTATGAAGTTCC 113
 Qy 301 CAGACTCTTCTCTACGATTTTGTGTCATATTTGGAATGCGTTTAACTTTCACCTTT 360
 Db 112 CAGACTCTTCTCTACGATTTTGTGTCATATTTGGAATGCGTTTAACTTTCACCTTT 53
 Qy 361 TAAATTTATGCTAACTTCTGATGAGTTCAATTAATATTTGACTAAAG 412
 Db 52 TAAATTTATGCTAACTTCTGATGAGTTCAATTAATATTTGACTAAAG 1

RESULT 10
 BP111459/c 683 bp mRNA linear EST 20-OCT-2000
 LOCUS BP111459
 DEFINITION 713412.x1 Soares NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:3523462.3 similar to TR:Q9UHP9 Q9UHP9 SMALL MUSCULAR
 PROTEIN. /, mRNA sequence.
 BP111459
 ACCESSION BP111459.1 GI:10941149
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40UP from G1bco
 High quality sequence stop: 470.
 Location/Qualifiers

FEATURES

source

1..683
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3523462"
 /clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and 88 circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clone IDs: Soares NBHP pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HP-9W pool 1:
 758280-760583, 772104-774407 Soares NBHPA pool 1:
 304776-306311, 320136-322823, 326280-326653 Soares NBHP
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima/Bonaldo."
 BASE COUNT 199 a 137 c 140 g 217 t
 ORIGIN

Query Match 90.5%; Score 382; DB 12; Length 683;
 Best Local Similarity 96.2%; Pred. NO. 2.2e-74;
 Matches 402; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
 Qy 1 GATGATGTGAAGAAATTAAGAGCAGAGATGATTCATAGCTCACTAAATTTTATA 60
 Db 461 GATGATGTGAAGAAATTAAGAGCAGAGATGATTCATAGCTCACTAAATTTTATA 402

Qy 61 TATTTGATGATGATTTGTGAACCTCTGAAATGCTGAGACTCTAGACAGAAATGGCTGTT 120
 Db 401 TATTTGATGATGATTTGTGAACCTCTGAAATGCTGAGACTCTAGACAGAAATGGCTGTT 342
 Qy 121 TGTACATTTATATCTCTTCTCTTCTAGTGGCTGTATTTCTTACTTTATCTTCATTTTGG 180
 Db 341 TGTACATTTATATCTCTTCTCTTCTAGTGGCTGTATTTCTTACTTTATCTTCATTTTGG 282
 Qy 181 CACCTCAGAGAACTTACCCCTAAATTCACACCTCGAGGCTGTGTTTGAAGAG 240
 Db 281 CACCTCAGAGAACTTACCCCTAAATTCACACCTCGAGGCTGTGTTTGAAGAG 222
 Qy 241 GATATGATTTTATGAGAAATGATATGCGCAATGCTCTTAACGATTTGAT-GAAAGTTTC 299
 Db 221 GATATGATTTTATGAGAAATGATATGCGCAATGCTCTTAACGATTTGATGCAAGTTTC 162
 Qy 300 CCAAGCTACTTCTCTACGATTTTGTGTCATATTTGGAATGCGTTTAACTTTCACCTT 359
 Db 161 CCAAGCTACTTCTCTACGATTTTGTGTCATATTTGGAATGCGTTTAACTTTCACCTT 102
 Qy 360 TTAATTTATGCTAACTTCTGATGAGTTCAATTAATATTTGACTAAATGTAATA 417
 Db 101 TTAATTTATGCTAACTTCTGATGAGTTCAATTAATATTTGACTAAATGTAATA 44

RESULT 11

F36800
 LOCUS F36800
 DEFINITION HSPD34697 HM3 Homo sapiens cDNA clone BHS-000020-0/B05, mRNA
 sequence.
 ACCESSION F36800.1 GI:4822426
 VERSION EST.
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 368)
 Lanfanchi, G., Muraro, T., Caldera, F., Pachioni, B., Pallavicini, A.,
 Pandolfo, D., Toppo, S., Trevisan, S., Scarzo, S. and Valle, G.
 Identification of 4370 expressed sequence tags from a
 3'-end-specific cDNA library of human skeletal muscle by DNA
 sequencing and filter hybridization
 Genome Res. 6 (1), 35-42 (1996)
 96276048
 JOURNAL MEDLINE
 COMMENT Contact: Valle G.
 CRIPI Biotechnology Centre
 University of Padua
 Via Trieste 75, 35121 Padua, Italy
 ABI Chromatograms and other information are available on WWW at
 http://grup.bio.unipd.it

FEATURES

source

1..368
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="BHS-000020-0/B05"
 /clone_id="HM3"
 /sex="Female"
 /tissue_type="pectoral muscle (after mastectomy)"
 /note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;
 Site 2: NotI; The library was constructed by G.
 Lanfanchi. This library is not subtracted nor normalized.
 The first strand cDNA was primed with a biotinylated
 oligo-dT-NotI primer
 (5'-biotin-AACCTCGCTCGAGCGCGCTTTTCTTTTCTTTT-3'). The
 ds cDNA was sonicated and size-selected in the range
 350-550 bp. The 3' specific fragments were selected by
 streptavidin coated magnetic beads, ligated to
 non-palindromic BstXI adapters, NotI digested and
 directionally cloned into BstXI-NotI cut pCDNAII vector."
 BASE COUNT 116 a 60 c 74 g 118 t
 ORIGIN

Query Match 74.9%; Score 316.2; DB 14; Length 368;
 Best Local Similarity 99.1%; Pred. No. 7.7e-60;
 Matches 318; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGGAGAGATGATTCATAGCTCACTAAATTTTATA 60
 DB 48 GATTGATGTGAAGAAATTAAGAGGAGATGATTCATAGCTCACTAAATTTTATA 107
 QY 61 TATTGATGTGAAGATGGAACCTCCGTAATCCGTAAGACCTGACAAATAGCGCTTT 120
 DB 108 TATTGATGTGAAGATGGAACCTCCGTAATCCGTAAGACCTGACAAATAGCGCTTT 167
 QY 121 TGTACATTTATATCTCTCTCTCTAGTGGCTGATTTCTTACTTTATCTTCATTTTGG 180
 DB 168 TGTACATTTATATCTCTCTCTCTAGTGGCTGATTTCTTACTTTATCTTCATTTTGG 227
 QY 181 CACCTCAGAGCAAAATTAAGCCCAATTAATCAACCTGAGGGGTGGTTTGAAGAG 240
 DB 228 CACCTCAGAGCAAAATTAAGCCCAATTAATCAACCTGAGGGGTGGTTTGAAGAG 287
 QY 241 GATATGATTTTATGAGAAATGATATGCAATGGCTTAACGATTTTGTGAAAAGTTTCC 300
 DB 288 GATATGATTTTATGAGAAATGATATGCAATGGCTTAACGATTTTGTGAAAAGTTTCC 347
 QY 301 CAAGCTACTTCTCTACAGTATT 321
 DB 348 CAAGCTACTTCTCTACAGTATT 368

RESULT 12
 AI824667/c 321 bp mRNA linear EST 17-DEC-1999
 LOCUS wc48c01.x1 NCI_CGAP_Pr28 Homo sapiens cDNA IMAGE:2321856 3,
 DEFINITION mRNA sequence.
 ACCESSION AI824667
 VERSION AI824667.1 GI:544538
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 321)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: CGAP@remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.livl.gov/bbrp/image/image.html
 Insert Length: 795 Std Error: 0.00
 Seq primer: -40UP from Gluco
 High quality sequence stop: 319.
 Location/Qualifiers
 1. .321

FEATURES
 source
 1. .321
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2321856"
 /clone_id="NCI_CGAP_Pr28"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI CGAP Pr22 was prepared, and as
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization

BASE COUNT 118 a 57 c 49 g 96 t 1 others

Query Match 71.3%; Score 300.8; DB 9; Length 321;
 Best Local Similarity 98.7%; Pred. No. 1.9e-56;
 Matches 313; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 107 AGAATGCGCTGTTTGTACA-TTATATCTCTCTCTAGTGGCTGATTTCTTACTT 165
 DB 321 AGAATGCGCTGTTTGTACA-TTATATCTCTCTCTAGTGGCTGATTTCTTACTT 262
 QY 166 TATCTCATTTTGGCAGCTCAGCAAGAAATTAAGCCCAATTAATCAACCTGAGGGT 225
 DB 261 TATCTCATTTTGGCAGCTCAGCAAGAAATTAAGCCCAATTAATCAACCTGAGGGT 202
 QY 226 GTGGTTTGAAGAGGATATGATTTTATGAGAAATGATGGAATGCTTAACGATTT 285
 DB 201 GTGGTTTGAAGAGGATATGATTTTATGAGAAATGATGGAATGCTTAACGATTT 142
 QY 286 TGATGAAAAGTTTCCCAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTT 345
 DB 141 TGATGAAAAGTTTCCCAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTT 82
 QY 346 TAGTTCTTCACTTTTAATTAATGTCATTAACCTTTGATGAGTCAATTAATTTTGA 405
 DB 81 TAGTTCTTCACTTTTAATTAATGTCATTAACCTTTGATGAGTCAATTAATTTTGA 22
 QY 406 CTAATGTAAATGTGA 422
 DB 21 CTAATGTAAATGTGA 5

RESULT 13
 F26916 313 bp mRNA linear EST 13-MAY-1999
 LOCUS HSPDI4595 HM3 Homo sapiens cDNA s400062D03, mRNA sequence.
 DEFINITION F26916
 ACCESSION F26916.1 GI:4812542
 VERSION EST.
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 313)
 Lanfranchi G., Muraro, T., Caldara, F., Pacchioni, B., Pallavicini, A.,
 Pandolfo, D., Toppo, S., Trevisan, S., Scaro, S. and Valle, G.
 Identification of 4370 expressed sequence tags from a
 3'-end-specific cDNA library of human skeletal muscle by DNA
 sequencing and filter hybridization
 Genome Res. 6 (1), 35-42 (1996)
 96276048
 COMMENT Contact: Valle G.
 CRIIBI Biotechnology Centre
 University of Padua
 Via Trieste 75, 35121 Padua, Italy
 ABI Chromatograms and other information are available on WWW at
<http://grup.bio.unipd.it>
 Location/Qualifiers
 1. .313
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="s400062D03"
 /clone_id="HM3"
 /sex="female"
 /tissue_type="pectoral muscle (after mastectomy)"
 /note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;
 Site 2: NotI; The library was constructed by G.
 Lanfranchi. This library is not subtracted nor normalized.
 The first strand cDNA was primed with a biotinylated

FEATURES
 source
 1. .313
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="s400062D03"
 /clone_id="HM3"
 /sex="female"
 /tissue_type="pectoral muscle (after mastectomy)"
 /note="Vector: pCDNAII (Invitrogen); Site 1: BstXI;
 Site 2: NotI; The library was constructed by G.
 Lanfranchi. This library is not subtracted nor normalized.
 The first strand cDNA was primed with a biotinylated

reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (cloneds
 985608-986759, 1101192-1101959, and 11217928-11220615).
 Subtraction by Bento Soares and M. Fatima Bonaldo. "

Query March	70.7%;	Score 298.4;	DB 14;	Length 313;
Best Local Similarity	99.7%;	Pred. NO. 6.5e-56;		
Matches 299;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTTGAATGCTCACTTAAATTTTATA 60

Db 14 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTTGAATGCTCACTTAAATTTTATA 73

QY 61 TATTTCATATGATATTTGTGAACCTCCTGAATGCTGAGACTCTAGACAAATGGCGCTGT 120

Db 74 TATTTCATATGATATTTGTGAACCTCCTGAATGCTGAGACTCTAGACAAATGGCGCTGT 133

QY 121 TGTACATTTATATCTCTCTCTCTAGTGGCGATATTTCTACTTATCTTCAATTTTGG 180

Db 134 TGTACATTTATATCTCTCTCTCTCTAGTGGCGATATTTCTACTTATCTTCAATTTTGG 193

QY 181 CACCTCAGAACCAATTAGCCCAATAATTCAACACTGAGGGTGTGTTTGAAGAG 240

Db 194 CACCTCAGAACCAATTAGCCCAATAATTCAACACTGAGGGTGTGTTTGAAGAG 253

QY 241 GATATGATTTTATGAGATGATATGAGCAATGTGCTTAAAGATTTTGAAGAAAGTTTC 300

Db 254 GATATGATTTTATGAGATGATATGAGCAATGTGCTTAAAGATTTTGAAGAAAGTTTC 313

RESULT	14
F31142	
LOCUS	F31142 302 bp mRNA EST 13-MAY-1999
DEFINITION	HSPD22104 HMJ Homo sapiens cDNA clone #400109D09, mRNA sequence.
ACCESSION	F31142
VERSION	F31142.1 GI:4816768
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumetazoa; Euthera; Primates; Catarrhini; Homnidae; Homo. 1 [bases 1 to 302]
AUTHORS	Lanfrench,I.G., Murro,T., Caldara,F., Pacchioli,B., Pallavicini,A., Pandofo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G. Identification of 4370 expressed sequence tags from a 3'-end-specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridization
TITLE	Genome Res. 6 (1), 35-42 (1996)
JOURNAL	
MEDLINE	96276048
COMMENT	Contact: Valle G.

FEATURES
source
1. .302
Location/Qualifiers
http://9rup.bwu.unipd.it.
ABI Chromatograms and other information are available on WWW at
Via Trieste 75, 35121 Padua, Italy
University of Padua
CIRI Biotechnology Centre

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="84000109009"
/clone_1b="HMJ3"
/sex="female"
/tissue_type="pectoral muscle (after mastectomy)"
/note="Vector: pCDNMI (Invitrogen); Site_1: BstXI;
Site_2: NotI; the library was constructed by G.
Lanfanchi. This library is not subtracted nor normalized
The first strand cDNA was primed with a biotinylated
oligo-dT-NotI primer

```

Query Match	69.1%	Score 291.4;	DB 14;	Length 302;
Best Local Similarity	99.7%;	Pred. No. 2.3e-54;		
Matches 292; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;

Qy	81	ACCTCCGATGATCGTAGACTGACCAAAATGSCGTGTGACATTTATATCTCTGC	140
Db	10	AATCTGAAATGCTGAGACTCTGACAAATATGCTGTGTGACATTTATATCTCTGC	69
Qy	141	TTCTATATGGCGATATTTCTTACTTATCTTCATTTTGGACCTCAGACCAATATAG	200
Db	70	TTCTATATGGCGATATTTCTTACTTATCTTCATTTTGGACCTCAGACCAATATAG	129
Qy	201	CCCATAAATTCAACACTGAGAGGATGTGTTTATAGAGGATATGTTTATATGGAAT	260
Db	130	CCCATAAATTCAACACTGAGAGGATGTGTTTATAGAGGATATGTTTATATGGAAT	189
Qy	261	GATATGGCAATGTGCTTACATTTTGTATGAAAATTTCCAAAGCTACTTCTCAAGAT	320
Db	190	GATATGGCAATGTGCTTACATTTTATATGAAAATTTCCAAAGCTACTTCTCAAGAT	249
Qy	321	TTTGGTCATAATTTGGAATGCGTTTATGTTCTTCACCTTTAATTAATATGAC	373
Db	250	TTTGGTCATAATTTGGAATGCGTTTATGTTCTTCACCTTTAATTAATATGAC	302

RESULT 15					
LOCUS	BM697544				
DEFINITION	BM697544	719 bp	mRNA	linear	EST 28-FEB-2002
ACCESSION	UI-B-DXO-sgn-1-12-0-UI r1 UI-B-DXO Homo sapiens CDNA clone				
VERSION	UI-B-DXO-sgn-1-12-0-UI 5', mRNA sequence.				
KEYWORDS	BM697544.1 GI:19010802				
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
JOURNAL	1 (bases 1 to 719)				
MEDLINE	Bonaldo,M.F., Lennon,G. and Soares,M.B.				
COMMENT	Normalization and subtraction: two approaches to facilitate gene genome Res. 6 (3), 791-806 (1996) 57044477				
Contact:	Soares, MB				

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mboares@iuiw.uweeg.uiaowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reegen.com).
Seq primer: M13 Reverse.

FEATURES
Location/Qualifiers
1. 719
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-DX0-agn-1-12-0-UI"
/clone_1b="UI-E-DX0"
/tissue_type="fetal eyes"

```

```

/dev stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-DX0 is a cDNA library containing the following
cist(s): fetal eyes. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tag for this library is AGAATCAGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
```

BASE COUNT 226 a 142 c 163 g 186 t 2 others

ORIGIN

Query Match 68.9%; Score 290.6; DB 14; Length 719;
 Best Local Similarity 98.0%; Pred. No. 3e-54;
 Matches 293; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

OY 1 GATTGATGTGAAGAAATTAAGAGGAGAGATGATTCATATAGCTCACTAAATTTTATA 60
    |||||||
DB 421 GATTGATGTGAAGAAATTAAGAGGAGAGATGATTCATATAGCTCACTAAATTTTATA 480
OY 61 TATTTGATGATGATTTGGAACCTCTGTAAGCTTAGAGCTGAGAAATGGCTGTT 120
    |||||||
DB 481 TATTTGATGATGATTTGGAACCTCTGTAAGCTTAGAGCTGAGAAATGGCTGTT 540
OY 121 TGTACATTTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
    |||||||
DB 541 TGTACATTTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
OY 181 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATATAGCTCACTAAATTTTATA 240
    |||||||
DB 601 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATATAGCTCACTAAATTTTATA 660
OY 241 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299
    |||||||
DB 661 GATATGATTTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719
```

Search completed: April 15, 2003, 22:02:37
 Job time : 877.253 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:26:41 ; Search time 112.61 Seconds
(without alignments) 8439.207 Million cell updates/sec

Title: US-09-647-019-10

Perfect score: 422
Sequence: 1 gactgactggaagaataaa.....tgactcaatgtaaatgta 422

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	422	100.0	886	24	AA27216
2	422	100.0	887	24	AA27216
3	412	97.6	587	24	AA27216
4	412	97.6	587	24	AA27216
5	146.8	34.8	778	20	AA27216
6	123	29.1	230	24	AA27216
7	60	14.2	60	24	AA27216
8	43	10.2	615	24	AA27216
9	41.2	9.8	6079	24	AA27216

10	41	9.7	13015	22	AA27216
11	41	9.7	13015	23	AA27216
12	41	9.7	13015	24	AA27216
13	39.8	9.4	3369	21	AA27216
14	39.8	9.4	3369	21	AA27216
15	39.6	9.4	1664976	19	AA27216
16	39.6	9.4	18154	24	AA27216
17	39.2	9.3	14672	19	AA27216
18	39.2	9.3	14672	19	AA27216
19	39.2	9.3	14672	19	AA27216
20	39.2	9.3	14672	19	AA27216
21	39.2	9.3	14672	19	AA27216
22	38.8	9.2	11718	22	AA27216
23	38.8	9.2	11718	22	AA27216
24	38.2	9.1	14383	23	AA27216
25	38.2	9.1	14383	23	AA27216
26	38.2	9.1	14383	23	AA27216
27	38.2	9.1	14383	23	AA27216
28	37.8	9.0	83391	24	AA27216
29	37.8	9.0	83391	24	AA27216
30	37.6	8.9	5236	24	AA27216
31	37.6	8.9	5236	24	AA27216
32	37.6	8.9	15416	24	AA27216
33	37.6	8.9	15416	24	AA27216
34	37.6	8.9	15416	24	AA27216
35	37.4	8.9	15416	24	AA27216
36	37.4	8.9	15416	24	AA27216
37	37.2	8.8	645	22	AA27216
38	37.2	8.8	645	22	AA27216
39	37.2	8.8	5439	24	AA27216
40	37.2	8.8	5439	24	AA27216
41	37.2	8.8	7490	24	AA27216
42	37.2	8.8	25002	19	AA27216
43	37	8.8	1766	22	AA27216
44	37	8.8	7445	24	AA27216
45	37	8.8	21537	24	AA27216

ALIGNMENTS

RESULT 1	AA27216	standard: DNA; 886 BP.
AC	AA27216	
DT	09-APR-2002	(first entry)
DE	Human 66214 EST clone DNA.	
KW	Human; congestive heart failure; dilative cardiomyopathy; sudden death; hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder; heart muscle disease; conduction disorder; coronary heart disease; systemic arterial hypertension; pulmonary hypertension; endocarditis; pulmonary heart disease; valvular heart disease; pericardial disease; congenital heart disease; gene therapy; syncope; transgenic animal; expressed sequence tag; EST; clone 66214; ds.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
XX	CDS	184..450
XX	FT	/*tag= a
XX	FT	/product= "Human 66214 protein"
XX	FT	298..588
XX	FT	/*tag= b
XX	FT	/note= "66214 cDNA fragment"
XX	FT	857..862
XX	FT	/*tag= c
XX	PN	MO200192567-A2.

Drosophila tetract
Drosophila melanog
Ampv genome fragm
Plasmodium falcipa
Methanococcus jann
Human immune syste
Streptococcus pneu
Human chemically p
Human immune syste
Human immune syste
Human immune/haema
Staphylococcus epi
Drosophila melanog
Mycobacteria genital
Human immune syste
Human angiotensin
Human polynucleoti
Human immune syste
Enterococcus faeca
Chemically treated
Human immune syste
Human gene regulat
Signal transductio
Human angiotensin
Human immune syste
Human breast cance
Streptococcus pneu
Human immune syste
Human immune syste
M. crenulata KHL
Human chemically p
Human immune syste
Human immune syste

PT muscular myopathies -
 XX
 XX Claim 8; Page 149-150; 157pp; English.
 CC
 CC The present sequence is the cDNA encoding the human caldesmon gene (Cald)
 CC that is mapped to the human chromosome Xq21.3-q22. It comprises 5 exons.
 CC It is a member of the EF-Hand protein super family and is involved in
 CC signaling pathways. It is predominantly expressed in heart and skeletal
 CC muscles and is activated after the differentiation of cells. Cal
 CC functions in regulation aspects of differentiation or adaptive processes
 CC that maintain muscle homeostasis. This sequence can be used in the
 CC detection, diagnosis, prophylactic and therapeutic treatment of diseases
 CC such as those involving aberrant muscle cell development and functional
 CC dysregulation, skeletal myopathies such as Duchenne muscular dystrophy and
 CC Becker's myotonic dystrophy, heart failure, cardiac hypertrophy,
 CC myocarditis, myofiber atrophy, etc. The Cal gene sequence can also be
 CC used in gene therapy for the production of transgenic animals and for
 CC drug screening.
 XX
 SQ Sequence 887 BP; 279 A; 172 C; 192 G; 244 T; 0 other;

Query Match 100.0%; Score 422; DB 20; Length 887;
 Best Local Similarity 100.0%; Pred. No. 1.7e-98;
 Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGGAAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 60
 DB 466 GATTGATGGAAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 525
 QY 61 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 526 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585
 QY 121 TGTACATTTATATCT 180
 DB 586 TGTACATTTATATCT 645
 QY 181 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 240
 DB 646 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 705
 QY 241 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 706 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 765
 QY 301 CAGCTACTTCT 360
 DB 766 CAGCTACTTCT 825
 QY 361 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 826 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 885
 QY 421 GA 422
 DB 886 GA 887

RESULT 3
 ABK64187/c
 ID ABK64187 standard; DNA; 587 BP.
 XX
 AC ABK64187;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human benign prostatic hyperplasia gene #82.
 XX
 KW Human, benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
 XX
 OS Homo sapiens.
 XX

PM WO200212440-A2.
 XX
 XX 14-FEB-2002.
 XX
 XX 07-AUG-2001; 2001MO-US24708.
 XX
 XX 07-AUG-2000; 2000US-223323P.
 XX
 XX 05-JUN-2001; 2001US-0873319.
 XX
 XX (GENE-) GENE LOGIC INC.
 PA (NISR) JAPAN TOBACCO INC.
 XX
 XX Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
 DR WPI; 2002-257476/30.
 XX
 PT Identifying drugs for and diagnosing benign prostatic hyperplasia, by
 PT detecting expression levels of one or more genes in prostate cells from
 PT patient that are differentially regulated compared to normal prostate
 PT cells -
 XX
 XX Disclosure; Page 104; 444pp; English.

The invention relates to a method of diagnosing (I) the onset or
 progression of benign prostatic hyperplasia (BPH), or screening (II) for
 or identifying an agent that modulates the onset or progression of BPH.
 The method is based on changes in gene expression in BPH tissue isolated
 from patients exhibiting different clinical states of prostate
 hyperplasia as compared to normal prostate tissue. (i) comprises
 detecting the expression levels of one or more genes in prostate cells
 from the subject that are differentially regulated compared to normal
 prostate cells. (ii) comprises preparing a first gene expression profile
 of BPH cells or BPH-like cell population, exposing the cells to the
 agent, preparing a second gene expression profile of the agent exposed
 cells, and comparing the first and second gene expression profiles.
 (1) is useful for identifying the onset or progression of BPH. (11) is
 useful for identifying an agent that modulates the onset or progression
 of BPH. The methods are useful to present information identifying
 the expression level in a tissue or cells, by comparing the expression
 level of genes given in the specification in the tissue or cells to the
 level of expression of gene in the database, and displaying the
 expression levels of at least one gene in the tissue or cells to the
 CC compared to the expression level in BPH. Agents using (11) are useful for
 CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
 CC benign prostatic hyperplasia gene sequences of the invention.
 XX
 SQ Sequence 587 BP; 177 A; 122 C; 103 G; 183 T; 2 other;

Query Match 97.6%; Score 412; DB 24; Length 587;
 Best Local Similarity 100.0%; Pred. No. 5.4e-96;
 Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGGAAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 60
 DB 412 GATTGATGGAAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 353
 QY 61 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 DB 352 TATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293
 QY 121 TGTACATTTATATCT 233
 DB 292 TGTACATTTATATCT 233
 QY 181 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 240
 DB 232 CACCTCAGAGAAATTAAGAGGAGAGATGATTCATAGTCACTAAATTTTATA 173
 QY 241 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 172 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 113
 QY 301 CAGCTACTTCT 360

DT 17-JAN-2000 (first entry)
 XX cDNA encoding murine chisel (Csl) gene.
 XX
 KM Chisel gene; Csl; EF-Hand protein super family; muscle development;
 KM heart/skeletal muscle cell development; signalling pathway; murine;
 KM X chromosome; regulation; adaptive process; muscle homeostasis;
 KM detection; diagnosis; prophylaxis; treatment; skeletal myopathy;
 KM muscular and myotonic dystrophy; Duchenne muscular dystrophy; therapy;
 KM Becker's myotonic dystrophy; heart failure; cardiac hypertrophy;
 KM differentiation; gene therapy; transgenic animal; drug screening; se.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 199..456
 FT /tag a
 FT /product "Chisel (Csl) protein"
 FT /note "Expressed especially in heart muscles"
 XX
 XX WO950410-A1.
 XX
 XX 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WO-AU00220.
 XX
 XX 27-MAR-1998; 99AU-0002634.
 XX
 XX (CHAN-) CHANG CARDIAC RES INST VICTOR.
 XX (GENO) GEN HOSPITAL CORP.
 XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
 XX
 XX Harvey RP, Muearo A, Palmer SJ, Rosenthal NA;
 XX WPI; 1999-610852/52.
 XX P-PSDB; AAY28650.
 XX
 XX Isolated nucleic acids encoding chisel, used to develop products for
 XX treating cardiomyopathy, cardiac hypertrophy, heart failure and
 XX muscular myopathies -
 XX
 XX Claim 3; Page 148; 157pp; English.
 XX
 CC The present sequence is the cDNA encoding the murine chisel gene (Csl)
 CC that is mapped to the mouse X chromosome. It is a member of the EF-Hand
 CC protein super family and is involved in signalling pathways. It is
 CC predominantly expressed in heart and skeletal muscles and is activated
 CC after the differentiation of cells. Csl functions in regulation aspects
 CC of differentiation or adaptive processes that maintain muscle
 CC homeostasis. This sequence can be used in the detection, diagnosis,
 CC prophylactic and therapeutic treatment of diseases such as those
 CC involving aberrant muscle cell development and functional activity. It
 CC is also used in the treatment of muscular and myotonic dystrophies,
 CC skeletal myopathies such as Duchenne muscular dystrophy and Becker's
 CC myotonic dystrophy, heart failure, cardiac hypertrophy, myocarditis,
 CC myofiber atrophy, etc. The Csl gene sequence can also be used in gene
 CC therapy, for the production of transgenic animals and for drug screening.
 CC
 SO Sequence 778 BP; 231 A; 166 C; 179 G; 202 T; 0 other;
 Query Match 34.8%; Score 146.8; DB 20; Length 778;
 Best Local Similarity 78.8%; Pred. No. 4,3e-28;
 Matches 175; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 OY 35 ATTGAATGCTCACTAAATTTTATATATTTGATGATGGAAGCTCTGGAATGCC 94
 DB 551 ATTAACAACCTCTGGAACATTTTATACATTTGTATGATGATCAACACCTCTGGAATGCC 610
 OY 95 TGAGACTCTAGCAGAAATGCGCTGTTGTACATTATATCTCTCTTCTAGTGGCTGT 154
 DB 611 CAGAGCTCTGCAAAATATCTCTTTGTCATTTATATTTCTCTTACTTGGTTC 670
 OY 155 ATTCTTACTTATCTTCAATTTTGGACCTCAGAAACAAATTAGCCATAATTCAC 214

DB 671 ATTTCTACCTTAGCTACATTTTGGACCTTGTAGAGCAATGACAGCAATTTACA 730
 OY 215 ACCTGAGAGGTGTGCTTTTGGAGAGGATATGATTTATGGA 256
 DB 731 ACCTGGAGAGGTGTGCTTTTGGAGAGGATATGATTTATGGA 772
 RESULT 6
 AAD27226
 ID AAD27226 standard; cDNA; 290 BP.
 XX
 AC AAD27226;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Human 66214 cDNA clone.
 XX
 KM Human; congestive heart failure; dilative cardiomyopathy; sudden death;
 KM hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
 KM heart muscle disease; conduction disorder; coronary heart disease;
 KM systemic arterial hypertension; pulmonary hypertension; endocarditis;
 KM pulmonary heart disease; valvular heart disease; pericardial disease;
 KM congenital heart disease; gene therapy; syncope; transgenic animal;
 KM clone 66214; se.
 XX
 OS Homo sapiens.
 XX
 XX WO200192567-A2.
 XX
 XX 06-DEC-2001.
 XX
 XX 30-MAY-2001; 2001WO-EP06165.
 XX
 XX 30-MAY-2000; 2000US-207400P.
 XX
 XX (MEDI-) MEDIGENE AG.
 XX
 XX Bunk D, Reuner B, Beck J, Henkel T;
 XX WPI; 2002-122073/16.
 XX
 XX Identifying a subject at risk for a heart disease e.g. congestive heart
 XX failure, dilative cardiomyopathy, heart muscle disease, by quantifying
 XX the polypeptide expressed by genes abnormally expressed in heart tissue
 XX
 XX Example 10; Fig 9a; 154pp; English.
 XX
 CC The patent discloses novel target genes abnormally expressed in heart
 CC tissues and their corresponding proteins. The invention also relates to
 CC methods for assessing the expression level of these genes. The method
 CC is used for testing the predisposition of mammals and preferably humans
 CC for a heart disease or for an acute state of such a disease. It is also
 CC useful to treat diseases of the heart such as congestive heart failure,
 CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
 CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
 CC syncope and sudden death, coronary heart disease, systemic arterial
 CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
 CC heart disease, congenital heart disease, pericardial disease and
 CC endocarditis. Sequences of the invention are also used in gene therapy.
 CC A transgenic non-human mammal comprising the sequences of the invention
 CC are useful for the development for medicaments for the treatments of
 CC heart diseases. The present sequence is a cDNA from 66214 clone. This
 CC sequence used in the exemplification of the invention.
 CC
 SO Sequence 290 BP; 103 A; 50 C; 69 G; 68 T; 0 other;
 Query Match 29.1%; Score 123; DB 24; Length 290;
 Best Local Similarity 100.0%; Pred. No. 4,2e-22;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GATTGATGTGAAGAAATTAAGAGGAGATGATTCATATAGCTCACTAAATTTTATA 60

DB 168 GATTGATCTGAAGAAATTAAGAGCAGAAATGATTCATAGCTCACTAAATTTTATA 227
|||
QY 61 TATTGTATGATGATTTGTGAACCTCTGGAATGCTGAGACTTTCGCGAANAATGGCCGTT 120
|||
DB 228 TATTGTATGATGATTTGTGAACCTCTGGAATGCTGAGACTTTCGCGAANAATGGCCGTT 287
|||
QY 121 TGT 123
|||
DB 288 TGT 290
|||
RESULT 7
ABN49389
ID ABN49389 standard; DNA; 60 BP.
XX
AC ABN49389;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human spliced transcript detection oligonucleotide SEQ ID NO:22137.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Homo sapiens.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-1B01903.
XX
PR 28-JUN-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Falgler S;
XX
DR WPI; 2002-257383/30.
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes -
XX
PS Example 1; SEQ ID 22137; 47bp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN72253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 60 BP; 16 A; 12 C; 14 G; 18 T; 0 other;
Query Match 14.2%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.8e-06;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 68 ATGATGATTTGTGAACCTCTGGAATGCTGAGACTTTCGCGAANAATGGCCGTTGTACAT 127
|||
DB 1 ATGATGATTTGTGAACCTCTGGAATGCTGAGACTTTCGCGAANAATGGCCGTTGTACAT 60
|||
RESULT 8
ABL33800
ID ABL33800 standard; DNA; 6115 BP.
XX
AC ABL33800;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1773.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 1773; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6115 BP; 1623 A; 47 C; 1519 G; 2926 T; 0 other;
Query Match 10.2%; Score 43; DB 24; Length 6115;
Best Local Similarity 59.5%; Pred. No. 0.26; 70; Indels 5; Gaps 2;
Matches 110; Conservative 0; Mismatches 0;
QY 228 AGGATGATGATTTTGTGAGATGATTTGCAATGCTTACGATTTTGAAGAAAGTT 297
|||
DB 5256 AGGATGATGATTTTGTGAGATGATTTGCAATGCTTACGATTTTGAAGAAAGTT 5313
|||

QY 298 TCCAGACTCTCTCTACAGATTTTGGTCAATATTTGGAATGGCTTTAGTCTCACC 357
 Db 5314 TTAGCTTTTATTTTATTTATTTTGGTAAATTTTGGTATTTTGGCTTTTATTT 5373
 QY 358 TTTAAATTAATGACCTAACTTTGTATGAGTTCAATTAATATTTGACCTAAATGTA 417
 Db 5374 TTTTATGTTATTTT---TAAAGTTTGTATTTATTAATAGGATTTTAAAGAGTAGA 5430
 QY 418 TGTGA 422
 Db 5431 AAGA 5435

RESULT 9

ABL32420
 ID ABL32420 standard; DNA; 6079 BP.

AC ABL32420;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 393.

KW Human; immune system disease; cytosine methylation; antineoplastic;
 KW antiarteriosclerotic; anti-HIV; anticonvulsant; ophthalmological;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineoplastic; antineoplastic; antidiabetic; antiparasitic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anemia;
 KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001MO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

PS Claim 1; SEQ ID NO 393; 32bp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/pulmonary bowel
 CC diseases. The present sequence is a gene of the invention.

SO Sequence 6079 BP; 1964 A; 94 C; 1241 G; 2780 T; 0 other;

Query Match

Best Local Similarity 9.8%; Score 41.2; DB 24; Length 6079;

Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 119 TTTGATATTATTAATCTCTCTCTAGTGGCTATTTCTACTTATCTTCATTTT 178
 Db 862 TTTGATATTATTTTGTATGAAGAGTTGATTAATTAATAGTATTAATTAATTTT 921

QY 179 GGCACCTCAGACAGAAATTAAGCCATTAATTCACACCTGAGGCTGTTGAGGA 238
 Db 922 TGTATTTAAATTAATTAATTTTAAATTTTAAATTAATTAATTAATTTGTTATGA 981
 QY 239 GGGATATATTTTATGAGAAATGATGCAATGCTGCTAAGATTTGATGAAGATT 298
 Db 982 TATGATATTTTATTAAGATGATGATTAATTAATTAATTTTAAATTTTATTAAT 1041
 QY 299 CCCAGCTACTCTCTACAGATTTTGGTCAATATTTGGAATG 340
 Db 1042 AAAATGCTATGATGAAGATGATTAATTAATTTGAAATG 1083

RESULT 10

AAF75571
 ID AAF75571 standard; DNA; 13015 BP.

AC AAF75571;

DT 09-MAY-2001 (first entry)

DE Drosophila tetratricopeptide repeat protein 2 regulatory region.

KW Animal model; polyglutamine tract; neurodegenerative disorder; HDJ;
 KW heat shock protein 40; tetratricopeptide repeat protein 2; TRP2; stroke;
 KW myeloid leukemia factor 1; MLF; human; fruit fly; Alzheimer's disease;
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma;
 KW cancer; ds.

OS Drosophila sp.

PN WO200112238-A1.

PD 22-FEB-2001.

PF 14-AUG-2000; 2000MO-US22496.

PR 12-AUG-1999; 99US-0148933.

PR 18-JAN-2000; 2000US-0177047.

PR 19-MAY-2000; 2000US-0205720.

PA (CALY) CALIFORNIA INST OF TECHNOLOGY.

PI Benzer S, Kazemi-Bafarjani P;

DR WPI; 2001-147537/15.

PT Identifying genes or other compounds that modulate polyglutamine
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease
 PT and Creutzfeldt-Jakob disease -

PS Disclosure; Fig 17; 275bp; English.

CC The present invention describes a method of screening for genes which
 CC modulate polyglutamine toxicity using animal models with polyglutamine
 CC sequences that cause toxicity in the animal. The model is preferably
 CC Drosophila, and toxic polyglutamine sequences include the human and
 CC (TRP2) and myeloid leukemia factor 1 (MLF) genes. The model is useful
 CC for identifying treatments for neurodegenerative and proliferative
 CC disorders, including Alzheimer's disease, Parkinson's disease,
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar
 CC ataxia, dentatorubropallidoluysian atrophy, Kennedy's disease, stroke,
 CC head trauma and cancer.

SO Sequence 13015 BP; 3746 A; 2643 C; 2515 G; 4111 T; 0 other;

Query Match

Best Local Similarity 9.7%; Score 41; DB 22; Length 13015;

Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 87 TGAATGCTGAGACTGAGAGAAATGCGCTGTTGATCATTTATCTTCCTCTAG 146
 DB 1103 TCAATAGACTGCAATATTTAGAAATTTATATATATTTGTTCTTCTTTT 1162
 QY 147 TTGGCTGATTTCTTACTTATCTTCATTTTGGCAGTCAAGAAATAGCCATA 206
 DB 1163 TTGGTTGGTTTTTTTTTTTTTTTTTTTTTTTCTTCTTGCAACCTTCCGCTCTC 1222
 QY 207 AATTCAACCTGAGGCTGTGTTTGGAGAG 239
 DB 1223 ATTTGACAGCCGAGAGTGGTGTTCAG 1255

RESULT 11
 ABL08614
 ID ABL08614 standard; cDNA; 21475 BP.
 XX
 AC ABL08614;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 20324.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;
 FI P-PEDB; ABB64511.
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 20324; 21pp + Sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB85737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 21475 BP; 6025 A; 4524 C; 4394 G; 6532 T; 0 other;
 SQ

Query Match 9.7%; Score 41; DB 23; Length 21475;
 Best Local Similarity 54.2%; Pred. No. 1.2;
 Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 87 TGAATGCTGAGACTGAGAGAAATGCGCTGTTGATCATTTATCTTCCTCTAG 146
 DB 1413 TCAATAGACTGCAATATTTAGAAATTTATATATATTTGTTCTTCTTTT 14192

QY 147 TTGGCTGATTTCTTACTTATCTTCAATTTTGGACCTGACAGAAATAGCCATA 206
 DB 1413 TTGGTTGGTTTTTTTTTTTTTTTTTTTTTTGTTCTTGCAACCTTCCGCTCTC 14252
 QY 207 AATTCAACCTGAGGCTGTGTTTGGAGAG 239
 DB 14253 ATTTGACAGCCGAGAGTGGTGTTCAG 14285

RESULT 12
 ABL56202/C
 ID ABL56202 standard; DNA; 50000 BP.
 XX
 AC ABL56202;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE ABL56202 genome fragment#4.
 XX
 KW ABL56202; gene therapy; viral vector; chromosome mapping; gene mapping;
 KW genetic deficiency disorder; ds.
 XX
 OS Amsacta moorei entomopoxvirus.
 XX
 PN WO200212526-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 10-AUG-2001; 2001WO-US25287.
 XX
 PR 10-AUG-2000; 2000US-224479P.
 XX
 PR 14-SEP-2000; 2000US-0662254.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Meyer RW, Li Y, Bawden AL;
 FI WPI; 2002-227161/28.
 XX
 DR Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence -
 XX
 PS Disclosure; Page 201-226; 326pp; English.
 XX
 CC The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
 CC (AmBPV).
 CC
 XX Sequence 50000 BP; 20248 A; 4709 C; 4703 G; 20340 T; 0 other;
 SQ

Query Match 9.7%; Score 41; DB 24; Length 50000;
 Best Local Similarity 52.7%; Pred. No. 1.4;
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 250 TTATGAGAAATGATGCAATGTCCTAACGATTTTGATGAGAAAGTTTCCCAAGCTACT 309
 DB 33101 TTTTGAAGATTTAAATGTTGTTTCATRTAGATGATATATATGAGCGAAGAAAT 33042
 OY 310 TCCACAGATATTTTGGTCAATATTTGGAAATCGTTTACGTTCTTACCTTTAAATTTATG 369
 DB 33041 TATTAATATTTAAATATATATATGATATATATATTTTACACACCGAATTTATC 32982
 OY 370 TCACATACTTGTATGAGTTCATATATATTTGACTTAATGTAAAT 418
 DB 32981 TAAATTATCTTTAAATGATTTATTAATGATTTTAAATTAATAAAT 32933

RESULT 13
 ID AAA70230/c
 AAA70230 standard; DNA; 3369 BP.

AC AAA70230;
 DT 07-NOV-2000 (first entry)
 DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:363.
 KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KM antimalarial; malaria; protozoa; infection; insecticide; ds.
 OS Plasmodium falciparum.
 PN WO200025728-A2.
 PD 11-MAY-2000.
 PF 05-NOV-1999; 99WO-US26796.
 PR 05-NOV-1998; 98US-0107131.
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 PI Hoffman S, Carnucci D, Gardner M, Venter JC;
 DR WPI; 2000-365347/31.
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 PS Disclosure; Page 550-551; 577pp; English.

CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasite lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB1814 to AAB1832 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

XX SQ Sequence 3369 BP; 1348 A; 425 C; 410 G; 1186 T; 0 other;
 Query Match 9.4%; Score 39.8; DB 21; Length 3369;
 Best Local Similarity 62.6%; Pred. No. 1.5;
 Matches 62; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
 OY 320 TTTTGGTCAATTTTGGAAATCGTTTACGTTCTTACCTTTAAATATATGACTAACT 379
 DB 537 TTTTATCCAAATTTTGTATTTTATATATATAGACATTTTATCTAATGATGATTAAG 478
 OY 380 TTGTATGATTCATATATATATTTGACTTAATGTAAAT 418
 DB 477 TTCAACATATCAACAATCTATACATCATGTAAAT 439

RESULT 14
 ID AAV21209/c
 AAV21209 standard; DNA; 1664976 BP.

AC AAV21209;
 DT 10-NOV-1998 (first entry)
 DE Methanococcus jannaschii circular chromosome.
 KM Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
 KM genome; autotrophic; extrachromosomal element; identification; ds.
 OS Methanococcus jannaschii.
 PN M09807830-A2.
 PD 26-FEB-1998.
 PF 22-AUG-1997; 97WO-US14900.
 PR 22-AUG-1996; 96US-0024428.
 PA (GENO-) INST GENOMIC RES.
 PA (UNIT) UNIV ILLINOIS FOUND.
 PA (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
 PI Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
 DR WPI; 1998-169145/15.
 PT Complete genome sequence of methano-genic archaeon, Methanococcus
 PT jannaschii - useful in identification of M. jannaschii genome
 PT fragment
 PS Claim 13; Page 152-585; 614pp; English.

CC The present sequence represents the complete 1.66-megabase pair genome
 CC sequence of the Methanococcus jannaschii circular chromosome. The
 CC present invention describes M. jannaschii open reading frames from the
 CC genome sequence. The invention also describes a computer based system
 CC for identifying fragments of the M. jannaschii genome that are
 CC homologous to target nucleotide sequences, comprising: (a) a data storage
 CC means comprising the nucleotide sequences of the 1664976, 58407 or 16550
 CC bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
 CC target sequence at least 99.9% identical to it; (b) search means for comparing a
 CC sequence to the nucleotide sequence of the data storage means to
 CC identify a homologous sequence, and (c) retrieval means for obtaining
 CC the homologous sequence. The method, which is based on whole genome
 CC random sequencing of an autotrophic archaeon M. jannaschii, the genome
 CC of which consists of 3 physically distinct elements, a large circular
 CC chromosome (the 1664976 bp sequence given in AAV21209), a large circular
 CC extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
 CC a small circular extra-chromosomal element (the 16550 bp sequence given
 CC in AAV21211), can be used in the identification of M. jannaschii genome
 CC fragment.

SQ Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;
 Query Match 9.4%; Score 39.8; DB 19; Length 1664976;
 Best Local Similarity 50.2%; Pred. No. 6.8;
 Matches 124; Conservative 0; Mismatches 122; Indels 1; Gaps 1;
 XX 114 GCCGTGTTGATCATTTATATCTCTCTAGTGGGTGATTTCTTACTTATCTTCA 173
 Db 1185113 GCTTTTTCCTCTTTAAAGCAGCTGCTATATATAAACAATATAGAAAAATCTTCA 1185054
 QY 174 TTTTGGACCTCAAGAACAAATTAAGCCATTAATTCACACCTGAGGGTGTGTTT 233
 Db 1185053 ACTTTCATATATCTTAATAATTCCTTACTGATTAAGCATTTGATGGATTTTAACT 1184994
 QY 234 GAGGAGGATATGATTTATGAGAAATATATGCAATGTCCTTAAGATTTTGAAT 292
 Db 1184993 GCACTAATTAAGTTATTTAATACAGAAATTAAGGAGTTTAAACATTAACAAAGAA 1184934
 QY 293 AAGTTCCCAAGCTACTTCTTACAGATTTTGGTCAATATTTGGAAATGCGTTTAACTTCT 352
 Db 1184933 CTGTTTATATGAAATTTCTTGGTATTTTGGTATTTACATTAATTAAGGTAAGATTTA 1184874
 QY 353 TCACCTT 359
 Db 1184873 GCTACTT 1184867
 RESULT 15
 ABL32254
 ID ABL32254 standard; DNA; 18154 BP.
 XX
 AC ABL32254;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 227.
 XX
 DE Human immune system associated gene SEQ ID NO: 227.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-amyloid; cytosine; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 227; 32bp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 18154 BP; 5274 A; 112 C; 3279 G; 9489 T; 0 other;
 Query Match 9.4%; Score 39.6; DB 24; Length 18154;
 Best Local Similarity 50.5%; Pred. No. 2.5;
 Matches 96; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
 QY 225 TGTGTTTGGAGAGGATATGATTTTATGAGAAATGATATGCAATGTGCTAAGATT 284
 Db 17300 TGTGATTTGGGATTTATATATTTTCTTCTGATTAATTTTATTTGTTAGATTT 17359
 QY 285 TTGATGAAAGTTTCCCAAGCTACTTCTTACAGATTTTGGTCAATTTTGAATGCTT 344
 Db 17360 TAGAGCTTGAGGGGAAATATATATTTGTTTGTGTTTATTTTTCGATTTT 17419
 QY 345 TTAGTTCTTCACTTTTAATATATATGCTTAACCTTGTATGAGTTCAATATATTTG 404
 Db 17420 TGTATTTTGGTTTGAAGATTTTATTTTATTTATTTAGTTGATTTGATTTT 17479
 QY 405 ACTAATGTA 414
 Db 17480 TTTAAAAAT 17489

Search completed: April 15, 2003, 19:13:32
 Job time : 626.61 secs

GenCore version 5.1.4 ps 4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 15, 2003, 16:47:06 ; Search time 956.114 Seconds

(without alignments)
12845.099 Million cell updates/sec

Title: US-09-647-019-10

Perfect score: 422

Sequence: 1 gactgactggaagaataaaa.....tgactcaaatgcaaatgtrga 422

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 205640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hhg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_pac:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vl:*
29: em_ba:*
30: em_fun:*
31: em_hum:*
32: em_in:*
33: em_mu:*
34: em_om:*
35: em_ov:*
36: em_pac:*
37: em_ph:*
38: em_pl:*
39: em_ro:*
40: em_sts:*
41: em_un:*
42: em_vl:*
43: em_ba:*
44: em_fun:*
45: em_hum:*

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	422	100.0	886	6	AX322774	AX322774 Sequence
2	422	100.0	886	6	AF129505	AF129505 Homo sapi
3	422	100.0	36503	9	U73508	U73508 Homo sapien
4	422	100.0	124186	9	AL772392	AL772392 Human DNA
5	421	99.8	885	9	HSR250584	HSR250584 Homo sapi
6	415.4	98.4	835	9	BC005948	BC005948 Homo sapi
7	412	97.6	587	6	AX322418	AX322418 Sequence
8	221.4	52.5	228031	2	AL772396	AL772396 Mus muscu
9	218	51.7	936	10	AF364070	AF364070 Mus muscu
10	211.4	50.1	943	10	MMU245772	MMU245772 Mus muscu
11	210.6	49.9	188670	2	AC127606	AC127606 Rattus no
12	210.4	49.9	188670	2	AF364071	AF364071 Rattus no
13	207.4	49.1	82586	2	AC096040	AC096040 Rattus no
14	146.8	34.8	787	10	AY026524	AY026524 Mus muscu
15	123	29.1	203251	2	AC124570	AC124570 Mus muscu
16	49.2	11.7	203251	2	AF288092	AF288092 Naegleria
17	46	10.9	49843	3	AF112367	AF112367 Plasmodiu
18	44.6	10.6	7622	3	AL157407	AL157407 Human DNA
19	44	10.4	142396	9	CEH02112	CEH02112 Homo sapi
20	43.2	10.2	34087	3	AC002531	AC002531 Homo sapi
21	43.2	10.2	197900	6	AX346702	AX346702 Sequence
22	43	10.2	6115	6	NM417719	NM417719 Nematode
23	42.8	10.1	13604	5	AL691516	AL691516 Zebraphish
24	42.8	10.1	80250	5	AP005437	AP005437 Oryza sat
25	42.6	10.1	73706	8	AC112837	AC112837 Rattus no
26	42.4	10.0	159651	2	AC116696	AC116696 Mus muscu
27	42.4	10.0	69730	2	AC129353	AC129353 Rattus no
28	41.8	9.9	113794	2	AC129353	AC129353 Rattus no
29	41.8	9.9	159952	2	AC007252	AC007252 Homo sapi
30	41.8	9.9	172034	2	AC125844	AC125844 Rattus no
31	41.8	9.9	184486	2	AC112579	AC112579 Rattus no
32	41.6	9.9	177067	2	AC016516	AC016516 Homo sapi
33	41.4	9.8	34465	3	U23520	U23520 Caenorhabdit
34	41.4	9.8	93463	8	ATP2206	ATP2206 Arabidops
35	41.4	9.8	129243	9	AC024603	AC024603 Homo sapi
36	41.2	9.8	6079	6	AX345322	AX345322 Sequence
37	41	9.7	46768	2	AC012871	AC012871 Drosophila
38	41	9.7	50000	6	AX392736	AX392736 Sequence
39	41	9.7	86004	9	HSR1J23	HSR1J23 Human DNA
40	41	9.7	169737	2	AC106594	AC106594 Rattus no
41	41	9.7	173704	2	AC126761	AC126761 Homo sapi
42	41	9.7	180479	2	AC129728	AC129728 Rattus no
43	41	9.7	232392	14	AF250284	AF250284 Ambascia m
44	41	9.7	254674	3	AE003651	AE003651 Drosophila
45	40.8	9.7	49820	9	AL353646	AL353646 Human DNA

ALIGNMENTS

RESULT 1
AX322774
LOCUS AX322774 886 bp DNA
DEFINITION Sequence 18 from Patent WO0192567.
ACCESSION AX322774
VERSION AX322774.1 GI:18093754
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
unclassified.

REFERENCE
1 Bunk, D., Reuner, B., Beck, J. and Henkel, T.
Novel target genes for diseases of the heart
Patent: WO 0192567-A 18 Dec-2001;
Medigene AG (DE)

FEATURES
Source
Location/Qualifiers
1..886
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 278 a 172 c 191 g 245 t

ORIGIN

Query Match 100.0%; Score 422; DB 6; Length 886;
Best Local Similarity 100.0%; Pred. No. 9.6e-79;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 60
DB 465 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 524

QY 61 TATTTGTATGATGATTTGGAACCTCTGAATGCCGAGACTGACAGAAATGGCTGTT 120
DB 525 TATTTGTATGATGATTTGGAACCTCTGAATGCCGAGACTGACAGAAATGGCTGTT 584

QY 121 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCAATTTTGG 180
DB 585 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCAATTTTGG 644

QY 181 CACCTCAGAGAACCAATTAAGCCCAATTAATTCACACCTGAGAGGTTGTTGAGAGG 240
DB 645 CACCTCAGAGAACCAATTAAGCCCAATTAATTCACACCTGAGAGGTTGTTGAGAGG 704

QY 241 GATATGATTTTATGAGATGATATGAGCAATGAGCTTAACGATTTTGAATGAAGTTTCC 300
DB 705 GATATGATTTTATGAGATGATATGAGCAATGAGCTTAACGATTTTGAATGAAGTTTCC 764

QY 301 CAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTTATGTTCTTCACTTT 360
DB 765 CAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTTATGTTCTTCACTTT 824

QY 361 TAAATATATGCTCAAACTTTGATGAGTTCAATTAATTTTGAATTAATGTAATTAAT 420
DB 825 TAAATATATGCTCAAACTTTGATGAGTTCAATTAATTTTGAATTAATGTAATTAAT 884

QY 421 GA 422
DB 885 GA 886

RESULT 2
AF129505 886 bp mRNA linear PRI 22-DEC-1999
LOCUS Homo sapiens small muscular protein (SMPX) mRNA, complete cds.
DEFINITION AF129505
ACCESSION AF129505.1 GI:6625646
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 886)
Patzak,D., Zhuchenko,O., Lee,C.C. and Wehner,M.
Identification, mapping, and genomic structure of a novel
X-chromosomal human gene (SMPX) encoding a small muscular protein
Hum. Genet. 105 (5), 506-512 (1999)

REFERENCE
AUTHORS
TITLE
JOURNAL
SOURCE
1..886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22.1-22.2"

gene
5'UTR
CDS
<1..>886
/gene="SMPX"
<1..183
/gene="SMPX"
184..450
/gene="SMPX"
/codon_start=1
/evidence="not experimental"
/product="small muscular protein"
/protein_id="AA19343.1"
/db_xref="GI:6625647"
/translation="MMSKOPVSNVRAIQANINIPMGAFRPAQOPPRRKECTPEVER
GVPSDEKXIPGAKLPGPAVLISIQIKSELKVPVPEQ"

misc_feature
3'UTR
polya_signal
BASE COUNT 279 a 172 c 191 g 244 t

ORIGIN

Query Match 100.0%; Score 422; DB 9; Length 886;
Best Local Similarity 100.0%; Pred. No. 9.6e-79;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 60
DB 465 GATTGATGTGAAGAAATTAAGAGGCGAAGATGATTCATAGCTCACTAAATTTTATA 524

QY 61 TATTTGTATGATGATTTGGAACCTCTGAATGCCGAGACTGACAGAAATGGCTGTT 120
DB 525 TATTTGTATGATGATTTGGAACCTCTGAATGCCGAGACTGACAGAAATGGCTGTT 584

QY 121 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCAATTTTGG 180
DB 585 TGTACATTTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTTATCTTCAATTTTGG 644

QY 181 CACCTCAGAGAACCAATTAAGCCCAATTAATTCACACCTGAGAGGTTGTTGAGAGG 240
DB 645 CACCTCAGAGAACCAATTAAGCCCAATTAATTCACACCTGAGAGGTTGTTGAGAGG 704

QY 241 GATATGATTTTATGAGATGATATGAGCAATGAGCTTAACGATTTTGAATGAAGTTTCC 300
DB 705 GATATGATTTTATGAGATGATATGAGCAATGAGCTTAACGATTTTGAATGAAGTTTCC 764

QY 301 CAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTTATGTTCTTCACTTT 360
DB 765 CAAGCTACTCTCTACAGTATTTTGGTCAATATTTGGAATGCGTTTATGTTCTTCACTTT 824

QY 361 TAAATATATGCTCAAACTTTGATGAGTTCAATTAATTTTGAATTAATGTAATTAAT 420
DB 825 TAAATATATGCTCAAACTTTGATGAGTTCAATTAATTTTGAATTAATGTAATTAAT 884

QY 421 GA 422
DB 885 GA 886

RESULT 3
U73508 36503 bp DNA linear PRI 27-APR-1999
LOCUS Homo sapiens cosmid clone U112E8 from Xp22.1-22.2, complete
DEFINITION U73508
ACCESSION U73508.1 GI:1616808
VERSION U73508.1
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 36503)

AUTHORS Sulston, J. E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998) X
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE The sequence of Homo sapiens cosmid clone U11288
 JOURNAL Unpublished (1999)
 REFERENCE 3 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submision
 JOURNAL Submitted (04-OCT-1996)
 REFERENCE 4 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submision
 JOURNAL Submitted (20-JAN-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 36503)
 AUTHORS Waterston, R.
 TITLE Direct Submision
 JOURNAL Submitted (27-APR-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT SUBMITTED BY: WUGSC
 Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:sapiens@waterston.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 This clone was mapped by Grief, M., Whyte, M. P., Thakker, R. V., and Mazzarella, R. Sequence analysis of 139 kb in xp22.1 containing spermine synthase and the 5' region of PEX. Genomics 44:227-231 (1997).

SOURCE INFORMATION:
 This clone is from a chromosome X-specific cosmid library LLOXNCC01 'U'. The source of the chromosomes was a human/hamster hybrid, GM07297-P, from Robert Nussbaum at the University of Pennsylvania School of Medicine. Please contact the Lawrence Livermore National Laboratory at http://www-bio.1lnl.gov/genome to obtain the clone.
 VECTOR: Lambda15C16.
 Location/Qualifiers
 1. 36503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="xp22.1-22.2"
 /clone="U11288"
 /clone_1lb="LLOXNCC01-U"
 116..196
 /rpt_family="L2"
 repeat_region
 1960..2046
 /rpt_family="L2"
 3690..4103
 /note="match to EST AA211443 (NID:g1810130) zn55D01.61"
 3711..3929
 /note="match to EST AA214031 (NID:g1812669) zn58F02.61"

misc_feature
 complement(3980..4103)
 /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
 repeat_region
 4141..4364
 /rpt_family="L1"
 repeat_region
 4366..4730
 /rpt_family="MaLR"
 repeat_region
 4732..4953
 /rpt_family="L1"
 repeat_region
 5586..5711
 /rpt_family="MaLR"
 repeat_region
 6433..6793
 /rpt_family="L1"
 repeat_region
 7169..7254
 /rpt_family="L2"
 repeat_region
 7874..8264
 /rpt_family="L1"
 repeat_region
 8529..8555
 /rpt_family="L2"
 repeat_region
 10938..11032
 /rpt_family="MIR"
 repeat_region
 11095..11460
 /rpt_family="MaLR"
 repeat_region
 11640..11719
 /rpt_family="L1"
 repeat_region
 12097..12191
 /rpt_family="L1"
 repeat_region
 12453..12595
 /rpt_family="L1"
 repeat_region
 13060..13222
 /rpt_family="MER1_type"
 repeat_region
 13285..13385
 /rpt_family="MER1_type"
 repeat_region
 13401..13625
 /rpt_family="MER1_type"
 repeat_region
 14107..14274
 /rpt_family="MIR"
 repeat_region
 14638..14751
 /rpt_family="MIR"
 repeat_region
 14804..15170
 /rpt_family="MaLR"
 repeat_region
 15277..15396
 /rpt_family="L2"
 repeat_region
 15512..15625
 /rpt_family="L2"
 repeat_region
 16941..17238
 /rpt_family="Alu"
 repeat_region
 17746..18251
 /rpt_family="MaLR"
 repeat_region
 19251..19321
 /rpt_family="MER1_type"
 repeat_region
 19385..19681
 /rpt_family="Alu"
 repeat_region
 20137..20192
 /rpt_family="L1"
 repeat_region
 20198..20256
 /rpt_family="MER1_type"
 repeat_region
 20456..20562
 /rpt_family="MIR"
 repeat_region
 20820..21010
 /rpt_family="MER1_type"
 repeat_region
 23777..23857
 /rpt_family="MIR"
 repeat_region
 24315..24471
 /rpt_family="MER1_type"
 repeat_region
 24480..24571
 /rpt_family="MIR"
 repeat_region
 24736..24842
 /rpt_family="L2"
 repeat_region
 26372..26608
 /rpt_family="MIR"
 repeat_region
 27011..27234
 /rpt_family="MER1_type"
 repeat_region
 28145..28286

```

repeat_region      /rpt_family="MIR"
                    28660..29064
                    /rpt_family="L1"
repeat_region      29526..29750
                    /rpt_family="L1"
repeat_region      30427..30777
                    /rpt_family="MADR"
repeat_region      30958..30988
                    /rpt_family="MER1_type"
repeat_region      34152..34277
                    /rpt_family="MIR"
repeat_region      34281..34328
                    /rpt_family="L2"
misc_feature        35221..35386
                    /note="match to EST AA211443 (NID:g1810130) zn55D01.s1"
misc_feature        complement(35221..35386)
                    /note="match to EST AA211521 (NID:g1810175) zn55D01.r1"
misc_feature        complement(35243..35386)
                    /note="match to EST AA389647 (NID:g2042633)"
repeat_region      36277..36501
                    /rpt_family="Retroviral"
BASE COUNT         10773 a 7354 c 7510 g 10866 t
ORIGIN
Query Match        100.0%; Score 422; DB 9; Length 36503;
Best Local Similarity 100.0%; Pred. No. 4e-79;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTGATGTGAAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 60
DB 4101 GATTGATGTGAAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 4042
QY 61 TATTGTATGATGATGATGTAACCTCTGTAATGCTGAGACTCTAGCAGAAATGCGCTGTT 120
DB 4041 TATTGTATGATGATGTAAGACCTCTGTAATGCTGAGACTCTAGCAGAAATGCGCTGTT 3982
QY 121 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 3981 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3922
QY 181 CACCTCAGAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 240
DB 3921 CACCTCAGAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 3862
QY 241 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 3861 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3802
QY 301 CAAGCTACTCTCTAGATTTTGTGCAATATTGGAATGCGTTTAACTTCTTCACTTT 360
DB 3801 CAAGCTACTCTCTAGATTTTGTGCAATATTGGAATGCGTTTAACTTCTTCACTTT 3742
QY 361 TAAATTAATGCTAAACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 3741 TAAATTAATGCTAAACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3682
QY 421 GA 422
DB 3681 GA 3680

```

```

REFERENCE
AUTHORS            1 (bases 1 to 124186)
                    Chapman,J.
TITLE              Direct Submission
JOURNAL            Submitted (18-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
                    Cambridgehire, CB10 1SA, UK. E-mail enquiries:
                    humquery@sanger.ac.uk Clone requests: cloneenquiries@sanger.ac.uk
                    On Jul 19, 2002 this sequence version replaced gi:21614755.
COMMENT
-----Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated with variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TrEMBL; Wp: WORMBEP; Information on the WORMBEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP11-450P7 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
Location/Qualifiers
source
1..124186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-450P7"
/clone_1lb="RP11-11.2"
BASE COUNT        37400 a 24595 c 23947 g 38244 t
ORIGIN
Query Match        100.0%; Score 422; DB 9; Length 124186;
Best Local Similarity 100.0%; Pred. No. 3e-79;
Matches 422; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATTGATGTGAAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 60
DB 100649 GATTGATGTGAAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 100590
QY 61 TATTGTATGATGATGTAACCTCTGTAATGCTGAGACTCTAGCAGAAATGCGCTGTT 120
DB 100589 TATTGTATGATGATGTAACCTCTGTAATGCTGAGACTCTAGCAGAAATGCGCTGTT 100530
QY 121 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
DB 100529 TGTACATTTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 100470
QY 181 CACCTCAGAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 240
DB 100469 CACCTCAGAGAAATTAAGAGGAGAGAGATGATGATCAATAGCTCACTAAATTTTATA 100410
QY 241 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 100409 GATATGATTTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 100350

```

QY 301 CAAGTACTTCTACAGATTTTGGTCAATATTTGGAAATGCGTTTAACTTCTTACCCCTT 360
 Db 100349 CAAGTACTTCTTACAGATTTTGGTCAATATTTGGAAATGCGTTTAACTTCTTACCCCTT 100290

QY 361 TAAATTATGTCATAACTTGTATGAGTCAATTAATATTTGACTAAATGCTAAATGCT 420
 Db 100289 TAAATTATGTCATAACTTGTATGAGTCAATTAATATTTGACTAAATGCTAAATGCT 100230

QY 421 GA 422
 Db 100229 GA 100228

RESULT 5
 HSA250584 885 bp mRNA linear PRI 12-APR-2001
 LOCUS Homo sapiens mRNA for stretch responsive muscle (X-chromosome)
 DEFINITION protein (Strm gene).
 ACCESSION AU250584.1 GI:10178976
 VERSION Strm gene; stretch responsive muscle (X-chromosome).
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 885)
 AUTHORS Kemp, T.J., Saduley, T.J., Simon, M., Brown, R., Eastwood, M., Sasseon, D.A. and Coulton, G.R.
 TITLE Identification of a novel stretch-responsive skeletal muscle gene (Smprx)
 JOURNAL Genomics 72 (3), 260-271 (2001)
 MEDLINE 21295047
 PUBMED 11401441
 REFERENCE 2 (bases 1 to 885)
 AUTHORS Kemp, T.J.
 TITLE Direct Submission
 JOURNAL Submitted (02-NOV-1999) Kemp T.J., Molecular pathology, Imperial College of Science Technology and Medicine, SAF Bldg, Level 2, Exhibition Road, South Kensington, London SW7 2AZ, UNITED KINGDOM
 COMMENT Related sequences: AJ245772, U73508 to U73509.
 FEATURES
 source location/Qualifiers
 1..885
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cfeature_type="skeletal muscle"
 1..885
 /gene="Strm"
 1..183
 /gene="Strm"
 184..450
 /gene="Strm"
 CDS
 5'UTR
 3'UTR
 poly_a_signal
 BASE COUNT 278 a 172 c 191 g 244 t
 ORIGIN
 Query Match 99.8%; Score 421; DB 9; Length 885;
 Best Local Similarity 100.0%; Pred. No. 1,6e-78;
 Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATTGATGTGAAGAAATTAAGAGCAGAGATGATCAATAGCTCAATAAATTTTATA 60
 Db 465 GATTGATGTGAAGAAATTAAGAGCAGAGATGATCAATAGCTCAATAAATTTTATA 524
 QY 61 TATTGTATGATGATTGTAACCTCTGAAATGCTGAGACTTACAGCAAGAAATGCGCTGTT 120

Db 525 TATTGTATGATGATTGTAACCTCTGAAATGCTGAGACTTACAGCAAGAAATGCGCTGTT 584
 QY 121 TGTACATTATATATCTCTCTCTTCTAGTGGCTGATTTCTTACTTTATCTTCATTTTGG 180
 Db 585 TGTACATTATATCTCTCTCTTCTAGTGGCTGATTTCTTACTTTATCTTCATTTTGG 644
 QY 181 CACCTCAGAGCAAAATTTAGCCCTAAATTCACACCTGAGGGGTGGTTGAGGAGG 240
 Db 645 CACCTCAGAGCAAAATTTAGCCCTAAATTCACACCTGAGGGGTGGTTGAGGAGG 704
 QY 241 GATATGATTTTATGAGAGATGATATGCGCAATGTCCTAACATTTTGGATGAAAGTTTCC 300
 Db 705 GATATGATTTTATGAGAGATGATATGCGCAATGTCCTAACATTTTGGATGAAAGTTTCC 764
 QY 301 CAAGTACTTCTTACAGATTTTGGTCAATATTTGGAAATGCGTTTAACTTCTTACCCCTT 360
 Db 765 CAAGTACTTCTTACAGATTTTGGTCAATATTTGGAAATGCGTTTAACTTCTTACCCCTT 824
 QY 361 TAAATTATGTCATAACTTGTATGAGTCAATTAATATTTGACTAAATGCTAAATGCT 420
 Db 825 TAAATTATGTCATAACTTGTATGAGTCAATTAATATTTGACTAAATGCTAAATGCT 884

QY 421 G 421
 Db 885 G 885

RESULT 6
 BC005948 835 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, small muscle protein, X-linked, clone MGC:14584
 DEFINITION IMAGE:4246501, mRNA, complete cds.
 ACCESSION BC005948
 VERSION BC005948.1 GI:13543590
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 835)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: CLONTECH
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-ehgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
 Clome distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.jnl.gov>
 Series: IRAL Plate: 21 Row: a Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6625646.
 Location/Qualifiers
 1..835
 /organism="Homo sapiens"
 /db_xref="LocusID:23676"
 /db_xref="taxon:9606"
 /clone="MGC:14584 IMAGE:4246501"
 /cfeature_type="Skeletal Muscle"
 /clone_1ib="NIH_MGC_81"

```

/lab_host="DH10B"
/ncbi="Vector: pDNR-LIB"
113..379
/codon_start=1
/product="small muscle protein, X-linked"
/db_xref="GI:13543591"
/translation="MNSKQPVSNVRAIQNNINPMGAFPPGAGQPPRRKCTPEVEE
GVPPTSEKKPIPGAKKLPQPAVNLSIENIKSELYKVKABQ"
BASE COUNT      282 a      155 c      171 g      227 t
ORIGIN

Query Match      98.4%; Score 415.4; DB 9; Length 935;
Best Local Similarity 99.8%; Pred. No. 2.4e-77;
Matches 416; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATTGATGTGAAGAAATAAGAGGAGAGATGATTCATAGCTCACTAAATTTTATA 60
DB 394 GATTGATGTGAAGAAATAAGAGGAGAGATTCATAGCTCACTAAATTTTATA 453
QY 61 TATTGTATGATGATGTGAAGCTCTGTAATGCTGAGACTCTAGCAGAAATGGCTGTT 120
DB 454 TATTGTATGATGATGTGAAGCTCTGTAATGCTGAGACTCTAGCAGAAATGGCTGTT 513
QY 121 TGTACATTTATATCTCTCTCTAGTTGGCTGTATTTCTTACTTATCTTCAATTTTGG 180
DB 514 TGTACATTTATATCTCTCTCTAGTTGGCTGTATTTCTTACTTATCTTCAATTTTGG 573
QY 181 CACCTCAGAGAACAAATTAAGCCCAATTAATTCACACCTGAGGGGTGGTTTGGAGAGG 240
DB 574 CACCTCAGAGAACAAATTAAGCCCAATTAATTCACACCTGAGGGGTGGTTTGGAGAGG 633
QY 241 GATATGATTTATGAGAGATGATATGCAATGTGCTTACATTTTGAATGAATTTTCC 300
DB 634 GATATGATTTATGAGAGATGATATGCAATGTGCTTACATTTTGAATGAATTTTCC 693
QY 301 CAAGTACTTCTCAGATTTTGTGCAATTTTGGATTTGCTTACTTCTTCACTTT 360
DB 694 CAAGTACTTCTCAGATTTTGTGCAATTTTGGATTTGCTTACTTCTTCACTTT 753
QY 361 TAAATTTATGCTAACTTTGTATGATGCTCAATTAATTTTGAATGAATGAATGA 417
DB 754 TAAATTTATGCTAACTTTGTATGATGCTCAATTAATTTTGAATGAATGAATGA 810

RESULT 7
AX332418 587 bp DNA linear PAT 09-JAN-2002
LOCUS      AX332418
DEFINITION Sequence 2927 from Patent WO0194629.
ACCESSION  AX332418
VERSION     AX332418.1 GI:18123052
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
             Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
             Horrikan,S., Soppet,D.R. and Weaver,Z.
             Cancer gene determination and therapeutic screening using signature
             gene sets
JOURNAL     Patent: WO 0194629-A 2927 13-DEC-2001;
             Avalon Pharmaceuticals (US)
             Location/Qualifiers
FEATURES
             source
             1..587
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
BASE COUNT  177 a      122 c      103 g      183 t      2 others
ORIGIN

Query Match      97.6%; Score 412; DB 6; Length 587;
Best Local Similarity 100.0%; Pred. No. 1.3e-76;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GATTGATGTGAAGAAATAAGAGGAGAGATGATTCATAGCTCACTAAATTTTATA 60
DB 412 GATTGATGTGAAGAAATAAGAGGAGAGATGATTCATAGCTCACTAAATTTTATA 353
QY 61 TATTGTATGATGATGTGAAGCTCTGTAATGCTGAGACTCTAGCAGAAATGGCTGTT 120
DB 352 TATTGTATGATGATGTGAAGCTCTGTAATGCTGAGACTCTAGCAGAAATGGCTGTT 293
QY 121 TGTACATTTATATCTCTCTCTAGTTGGCTGTATTTCTTACTTATCTTCAATTTTGG 180
DB 292 TGTACATTTATATCTCTCTCTAGTTGGCTGTATTTCTTACTTATCTTCAATTTTGG 233
QY 181 CACCTCAGAGAACAAATTAAGCCCAATTAATTCACACCTGAGGGGTGGTTTGGAGAGG 240
DB 232 CACCTCAGAGAACAAATTAAGCCCAATTAATTCACACCTGAGGGGTGGTTTGGAGAGG 173
QY 241 GATATGATTTATGAGAGATGATATGCAATGTGCTTACATTTTGAATGAATTTTCC 300
DB 172 GATATGATTTATGAGAGATGATATGCAATGTGCTTACATTTTGAATGAATTTTCC 113
QY 301 CAAGTACTTCTCAGATTTTGTGCAATTTTGGATTTGCTTACTTCTTCACTTT 360
DB 112 CAAGTACTTCTCAGATTTTGTGCAATTTTGGATTTGCTTACTTCTTCACTTT 53
QY 361 TAAATTTATGCTAACTTTGTATGATGCTCAATTAATTTTGAATGAATGAATGA 412
DB 52 TAAATTTATGCTAACTTTGTATGATGCTCAATTAATTTTGAATGAATGAATGA 1

RESULT 8
AL732396 228031 bp DNA linear HTG 17-AUG-2002
LOCUS      AL732396
DEFINITION Mus musculus chromosome X clone RP23-93M14, *** SEQUENCING IN
ACCESSION  AL732396
VERSION     AL732396.6 GI:22415936
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
             house mouse.
ORGANISM    Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
REFERENCE   1
             Heath,P.
             Direct Submission
             Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Aug 21, 2002 this sequence version replaced gi:22204493.
COMMENT     ----- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ----- Project Information
             Center project name: BM93M14
             ----- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 99% of reads
             Chemistry: Dye-terminator Big Dye; 0% of reads
             Consensus quality: 224691 bases at least Q40
             Consensus quality: 226956 bases at least Q30
             Insert size: 227331; sum-of-contigs
             Insert size: 194478; 8.5% error; agarose-fp
             Quality coverage: 6.42x in Q20 bases; sum-of-contigs Quality
             coverage: 8.63x in Q20 bases; agarose-fp

```

* NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 5544: contig of 5544 bp in length
 * 5545 5644: gap of 100 bp
 * 5645 11121: contig of 5477 bp in length
 * 11122 11221: gap of 100 bp
 * 11222 23110: contig of 11889 bp in length
 * 23111 23210: gap of 100 bp
 * 23211 42146: contig of 18936 bp in length
 * 42147 42246: gap of 100 bp
 * 42247 74925: contig of 32679 bp in length
 * 74926 75025: gap of 100 bp
 * 75026 153828: contig of 78803 bp in length
 * 153829 153928: gap of 100 bp
 * 153929 172635: contig of 18707 bp in length
 * 172636 172735: gap of 100 bp
 * 172736 228031: contig of 55296 bp in length.
 * Location/Qualifiers
 1. 228031
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-93M14"
 /clone_1fb="RPCT-23"
 1. 5544
 /note="assembly_fragment:01721
 fragment_chain:1"
 5645. 11121
 /note="assembly_fragment:02676
 fragment_chain:1"
 11222. 23110
 /note="assembly_fragment:00299
 fragment_chain:1"
 23211. 42146
 /note="assembly_fragment:01463
 fragment_chain:1"
 42247. 74925
 /note="assembly_fragment:02519
 fragment_chain:1"
 75026. 153828
 /note="assembly_fragment:00456
 fragment_chain:1"
 153829. 172635
 /note="assembly_fragment:03454
 fragment_chain:2"
 172736. 228031
 /note="assembly_fragment:01243
 fragment_chain:2"

BASE COUNT 71213 a 42522 c 42780 g 70814 t 702 others
 ORIGIN
 Query Match 52.5%; Score 221.4; DB 2; Length 228031;
 Best Local Similarity 78.3%; Pred. No. 2.9e-37;
 Matches 306; Conservative 0; Mismatches 76; Indels 9; Gaps 3;
 Oy 35 ATTCAATAGCTCACTAAATTTATATATTTGATGATGATGGAACCTCGTAATGCC 94
 Db 187401 ATAAACAACCTCTGACATTTTATACATTTGTATGATGATGACAAACCTCTGAATGCC 187460
 Oy 95 TGAGACTCTAGCAAAATGGCGCTGTGTATACATTTATATCTCTTCCTTACGTGGCTGT 154
 Db 187461 CAAGACTCTAGCAAAATATCTGTGTGTATATATTTCTTCCTTACCTGTGGC 187520
 Oy 155 ATTCTTACTTATCTTCAATTTTGGCACTCAGCAAAATATAGCCCAATATTCAC 214
 Db 187521 ATTCTCACTTATGCTATTTTGGCACTTGTAGCAAAATCAGCAGCAATTTACA 187580
 Oy 215 ACCTGGAGGGTGTGTTTGGAGAGGATATGA-TTTTATGAGAAATGATATGGCAATGT 273
 Db 187581 ACCTGGAGAGGTGTGTTTGGAGAGATGATTTTATGAAGGGGGGATGGCAATGT 187640
 Oy 274 GC--CTAAGATTTTGAATGAAGATTTCCCAAGCTACTCTCAAGATATTTTGTGTCATA 331

Db 187641 GCAAGCACTGATTTTGTATGATTTAGTACTTATAGTACTTCCACGGCTTTTGTGTCATA 187700
 Oy 332 TTGGAAAGCGCTTTTAGTCTTCACTTTTAAATATGATGATCACTTAACCTTGTATGATTC 391
 Db 187701 TTGGAAAG-----GTTTCTTACCTTTTAAATATATCTCAATTAACCTTTTATGATTC 187754
 Oy 392 AAATTAATATTTGACTAATATGTAATAATGTGA 422
 Db 187755 AAATTAATATTTGAGTAATATGTAACATATGA 187785

RESULT 9
 AF364070 936 bp mRNA linear ROD 04-MAY-2001
 LOCUS AF364070
 DEFINITION Mus musculus SMPX protein (Smpx) mRNA, complete cds.
 VERSION AF364070.1 GI:13940507
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 936)
 Patzak,D., Zhuchenko,O., Lee,C.C. and Wehnert,M.
 Identification, mapping, and genomic structure of a novel
 X-chromosomal human gene (SMPX) encoding a small muscular protein
 Hum. Genet. 105 (5), 506-512 (1999)
 MEDLINE 20065879
 PUBMED 10598820
 REFERENCE Patzak,D.
 2 (bases 1 to 936)
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-2001) E.M.-Arndt-Uni, Inst. f. Humang.,
 Fleischmannstr. 42-44, D-17487 Greifswald, Germany
 FEATURES
 1. 936
 Location/Qualifiers
 source
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 <1. 936
 /gene="Smpx"
 199. 456
 /gene="Smpx"
 /codon_start=1
 /product="SMPX protein"
 /protein_id="AAK50398.1"
 /db_xref="GI:13940508"
 /translation="MSKQPIISNVRAIQININIPMGAPRPGAGPPRRKESTPETEGE
 PTTSEKKPIQPMKKFPGPVNVLRIQVNSKLVKVPAGEQ"
 906. 911
 /gene="Smpx"
 /evidence=experimental

BASE COUNT 283 a 187 c 203 g 263 t
 ORIGIN
 Query Match 51.7%; Score 218; DB 10; Length 936;
 Best Local Similarity 78.2%; Pred. No. 5.5e-36;
 Matches 302; Conservative 0; Mismatches 75; Indels 9; Gaps 3;
 Oy 35 ATTCAATAGCTCACTAAATTTTATATTTTATGATGATGTAAGCTCGTAATGCC 94
 Db 551 ATAAACAACCTCTGACATTTTATACATTTGTATGATGATGACAAACCTCTGAATGCC 610
 Oy 95 TGAGACTCTAGCAAAATGGCGCTGTGTATACATTTATATCTCTTCCTTACGTGGCTGT 154
 Db 611 CAAGACTCTAGCAAAATATCTGTGTGTATATATTTCTTCCTTATCTTGTGGC 670
 Oy 155 ATTCTTACTTATCTTCAATTTTGGCACTCAGCAAAATATAGCCCAATATTCAC 214
 Db 671 ATTCTCACTTATGCTATTTTGGCACTTGTAGCAAAATCAGCAGCAATTTACC 730
 Oy 215 ACCTGGAGGGTGTGTTTGGAGAGGATATGA-TTTTATGAGAAATGATATGGCAATGT 273

misc_feature	740..744	/gene="Srmx"	/function="mRNA destabilising motif"
misc_feature	779..785	/gene="Srmx"	/function="mRNA destabilising motif"
misc_feature	909..913	/gene="Srmx"	/function="mRNA destabilising motif"
polyA_signal	923..928	/gene="Srmx"	/function="mRNA destabilising motif"
polyA_site	943	/gene="Srmx"	
BASE COUNT	278 a	192 c	210 g 262 t 1 others
ORIGIN			
Query Match	50.1%;	Score 211.4;	DB 10; Length 943;
Best Local Similarity	77.5%;	Pred. No. 1.3e-34;	
Matches 297; Conservative	0;	Mismatches 76;	Indels 10; Gaps 3;
Qy	35	ATTCAATAGCTCACTAAATTTATATATTTGATATGATATGTAATGGAACCTCTGGAATGCC	94
Db	567	ATTAACAACATCTCCGAACATTTTATACATTTGATATGATATGATCAACAACTCTCGAATGCC	626
Qy	95	TGAACTCTGACAGAAATGCGCTGTTGTCATTAATATATATGCTCTGCTGAGTGCGCTGT	154
Db	627	CAAACTCTGACAAATATATCTGTTGACATTTATATATTTCTTCTTTACTTGTTGC	686
Qy	155	ATTCTTACTTTATCTTCATTTTGGCACCCTGACAGAACAAATTAGCCATTAATTCAC	214
Db	687	ATTCTCACTTTACTACATATTTTGGCACCCTGAGAGCAAAATCAGACACGAATTTTACA	746
Qy*	215	ACCTGAGAGGCTGCTTTGAGAGGAGATATGATTTATGA--GAATGATATGGCAATG	272
Db	747	ACCTGGAGAGGCTGCTTTGAGAGAGATGATTTTATGAAGGGGGAGATGGCAACG	806
Qy	273	TGC--CTAACGATTTTATGATGAAGTTTCCCAAGCTACTTCTCAAGTATTTGGTCAAT	330
Db	807	TGCAGACGATGATTTTATGATGTTAAGTACTTTAAGTTCTCCACGGCTTTGGTCAAT	866
Qy	331	ATTGGAATGCGTTTATGTTCTTCACTTTAAATATATATGTCATTAACCTTGTATGATT	390
Db	867	ATTGGAATG-----GTTCTTCACTTTTAAATATCTCAATTAACCTTTTATGAGTT	920
Qy	391	CAATTAATATTTGACTAATATG	413
Db	921	CAATTAATATTTGAGTAAATGT	943
RESULT 11			
AC127606	188670 bp	DNA	linear HTG 18-JUL-2002
LOCUS			
DEFINITION	Rattus norvegicus clone G230-20D15, *** SEQUENCING IN PROGRESS		
ACCESSION	AC127606		
VERSION	AC127606.1	GI:21902814	
SOURCE	HTGS PHASE1.		
ORGANISM	Normay rat.		
	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 188670)		
AUTHORS	Murny,D.M., Adams,C., Adio-Oduola,B., Ali-oesman,F.R., Allen,C., Albrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbatoria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Burke,C., Burrell,K.L., Byrd,N.C., Bunay,C., Burck,P., Burkett,C., Cavazos,S.R., Chacko,V., Chavez,D., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,V., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dedereich,D.A., Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,		

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Barnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frenz, P.,
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B.,
Hornel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsberg, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, M., Loubege, H.,
Lorado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maneshwar, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moore, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokkenko, S., Ogun, W., Okunodu, G.,
Orzguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoohart, N., Sison, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Swatek, A., Tabor, P., Tamerias, A., Tamerias, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Teitrod, B., Thomas, N., Thomas, S.,
Umanan, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watling, S.,
Williams, G., Williamson, A., Wlarczyk, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. P., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Unpublished
2 (bases 1 to 188670)

Worley, K. C.
Submitted (18-JUL-2002)

Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc.helpebcm.tmc.edu

Project Information

Center project name: GX0D

Center clone name: CH230-20D15

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye 1000 of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 142614 bases at least Q40

Consensus quality: 149053 bases at least Q30

Consensus quality: 154556 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).
NOTE: This is a 'working draft' sequence. It currently
consists of 60 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1104: contig of 1104 bp in length
1204: gap of unknown length
1205: contig of 1287 bp in length
2492: gap of unknown length
2591: gap of unknown length
3744: contig of 1153 bp in length
3844: gap of unknown length
3845: contig of 1580 bp in length
5425: gap of unknown length
5525: contig of 1324 bp in length

6848: gap of unknown length

6849: gap of unknown length
6949: contig of 1272 bp in length
8220: gap of unknown length
8221: contig of 1298 bp in length
8321: gap of unknown length
9619: contig of 1261 bp in length
9718: gap of unknown length
10979: contig of 1110 bp in length
11080: gap of unknown length
12180: contig of 1507 bp in length
12280: gap of unknown length
13797: contig of 1506 bp in length
13896: gap of unknown length
15402: contig of 1194 bp in length
15502: gap of unknown length
15503: contig of 1595 bp in length
16597: gap of unknown length
16797: contig of 1595 bp in length
18392: gap of unknown length
18492: contig of 2358 bp in length
20850: gap of unknown length
20950: contig of 2098 bp in length
23048: gap of unknown length
23148: contig of 2487 bp in length
25635: gap of unknown length
25735: contig of 1155 bp in length
26889: gap of unknown length
26990: contig of 2246 bp in length
29235: gap of unknown length
29335: contig of 1470 bp in length
29336: gap of unknown length
30805: contig of 1320 bp in length
30906: gap of unknown length
32225: contig of 1982 bp in length
32326: gap of unknown length
32325: contig of 1535 bp in length
34407: gap of unknown length
34408: contig of 1535 bp in length
35942: gap of unknown length
35943: contig of 2075 bp in length
36043: gap of unknown length
38217: contig of 1985 bp in length
38318: gap of unknown length
40302: contig of 2222 bp in length
40303: gap of unknown length
42524: contig of 1896 bp in length
42525: gap of unknown length
44520: contig of 3091 bp in length
44521: gap of unknown length
47711: contig of 2351 bp in length
47811: gap of unknown length
47812: contig of 1981 bp in length
50163: gap of unknown length
50262: contig of 1981 bp in length
52243: gap of unknown length
52244: contig of 2805 bp in length
55148: gap of unknown length
55248: contig of 2731 bp in length
57979: gap of unknown length
58079: contig of 2318 bp in length
58080: gap of unknown length
60397: contig of 2523 bp in length
60398: gap of unknown length
63020: contig of 3893 bp in length
63120: gap of unknown length
63121: contig of 1957 bp in length
67014: gap of unknown length
67114: contig of 1957 bp in length
69070: gap of unknown length
69071: contig of 2226 bp in length
71396: gap of unknown length
71397: contig of 2093 bp in length
71496: gap of unknown length
73589: contig of 2949 bp in length
73590: gap of unknown length
76638: contig of 2877 bp in length
76639: gap of unknown length
79615: contig of 3028 bp in length
79616: gap of unknown length
82744: contig of 3421 bp in length
82745: gap of unknown length
86264: contig of 3421 bp in length
86265: gap of unknown length

RESULT 13
AC096040/c
LOCUS
DEFINITION AC096040 82586 bp DNA linear HTG 11-JUL-2002
46 unordered pieces.
AC096040
AC096040.4 GI:21723170
HTG; HTGS PHASE1.
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 82586)

Wuzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooke,S.L., Amarutunge,H.C., Are,J.R., Ayalew,M., Banks,T.,
Barberia,J., Benton,J., Blumig,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Doutwaite,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Dunham,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homel,P., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudh,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kraevic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,M., Louised,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapue,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mathew,E., McLeod,M.P., Meador,M., Mei,G., Metker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moher,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Ogun,M., Okunolu,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Sodergren,S., Scott,G., Shen,H., Shoshitani,N., Sisson,I.,
Sutton,A., Swalek,A., Taber,P., Tamerias,A., Tamerias,K., Tang,H.,
Tantony,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vazquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinbeck,G., and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

COMMENT

Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:17943701.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GSGR
Center clone name: CH230-2911
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 35907 bases at least Q40
Consensus quality: 40324 bases at least Q20
Consensus quality: 44113 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 46 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1317:	contig of 1317 bp in length
1318	1417:	gap of unknown length
1418	2800:	contig of 1383 bp in length
2801	2900:	gap of unknown length
2901	4244:	contig of 1344 bp in length
4245	4344:	gap of unknown length
4345	5864:	contig of 1520 bp in length
5865	5964:	gap of unknown length
5965	6968:	contig of 1004 bp in length
6969	7068:	gap of unknown length
7069	8166:	contig of 1098 bp in length
8167	8266:	gap of unknown length
8267	9384:	contig of 1118 bp in length
9385	9484:	gap of unknown length
9485	10630:	contig of 1146 bp in length
10631	10730:	gap of unknown length
10731	11858:	contig of 1128 bp in length
11859	11958:	gap of unknown length
11959	13413:	contig of 1455 bp in length
13414	13513:	gap of unknown length
13514	14847:	contig of 1334 bp in length
14848	14947:	gap of unknown length
14948	16180:	contig of 1233 bp in length
16181	17804:	gap of unknown length
17805	17904:	contig of 1524 bp in length
17905	18913:	gap of unknown length
18914	19013:	contig of 1009 bp in length
19014	20050:	contig of 1037 bp in length
20051	20150:	gap of unknown length
20151	21378:	contig of 1228 bp in length
21379	21478:	gap of unknown length
21479	22828:	contig of 1350 bp in length
22829	22928:	gap of unknown length
22929	23998:	contig of 1070 bp in length
23999	24098:	gap of unknown length
24099	25596:	contig of 1498 bp in length
25597	26565:	gap of unknown length
26566	27065:	contig of 1269 bp in length
27066	29291:	gap of unknown length
29292	29391:	contig of 2226 bp in length
29392	30870:	gap of unknown length
30871	30970:	contig of 1479 bp in length
30971	33590:	gap of unknown length
33591	33690:	contig of 2620 bp in length
33691	34829:	gap of unknown length
34830	34928:	contig of 1139 bp in length
34930	36000:	gap of unknown length
36001	36700:	gap of unknown length

QY 215 ACCTGAGAGGTGCTGTTTGAGAGGATGATTTATGGA 256
 Db 738 ACCTGGGAGAGTGTGTTTGAGAGAGATGATTTATGA 779

RESULT 15

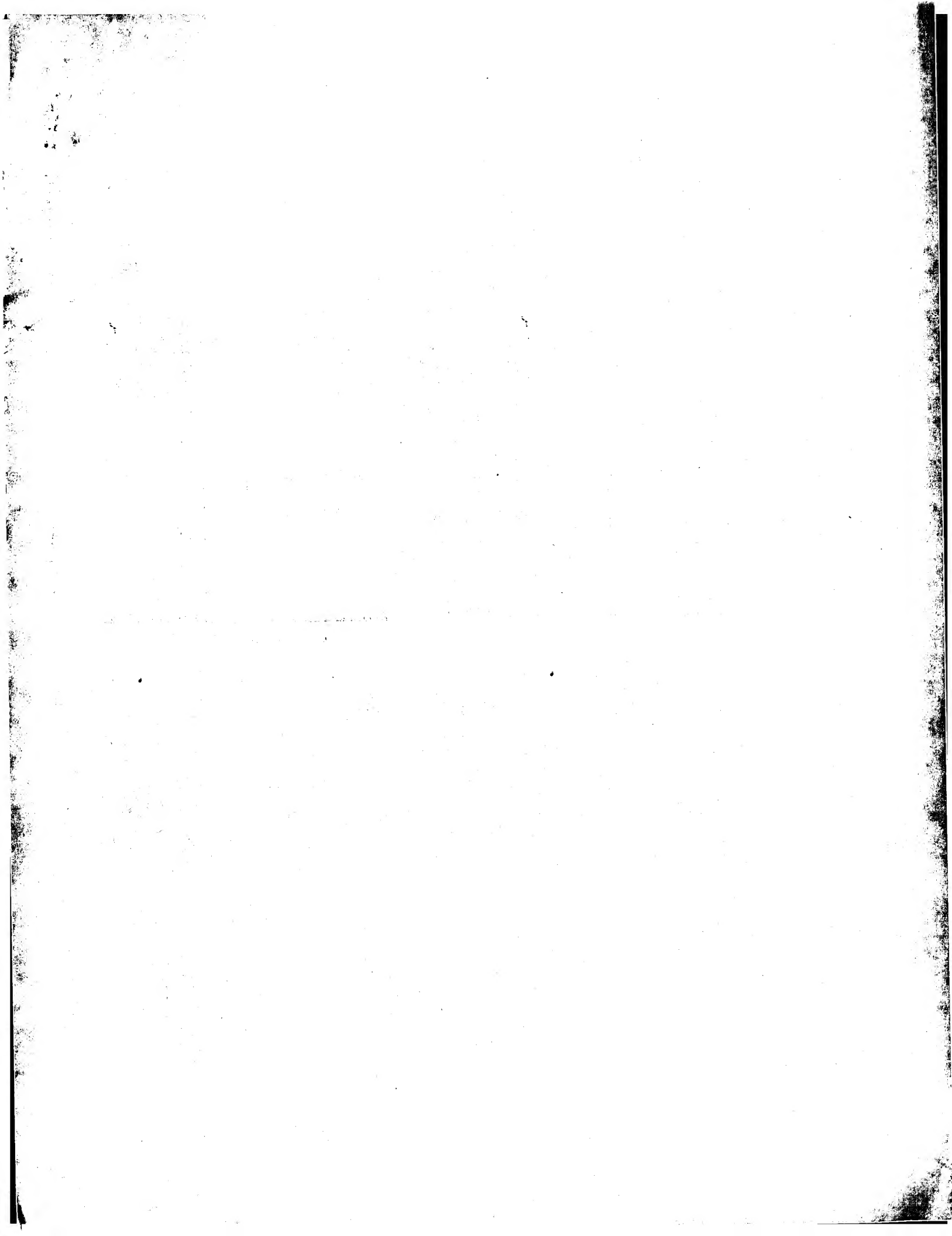
AX322783
 LOCUS AX322783 290 bp DNA linear PAT 07-JAN-2002
 DEFINITION Sequence 27 from Patent WO0192567.
 ACCESSION AX322783
 VERSION AX322783.1 GI:18093762
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 unidentified.
 1 Bunk,D., Reuner,B., Beck,J. and Henkel,T.
 TITLE Novel target genes for diseases of the heart
 JOURNAL Patent: WO 0192567-A 27 06-DEC-2001;
 Medigene AG (DE)
 FEATURES
 source Location/Qualifiers
 1..290
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 103 a 50 c 69 g 68 t
 ORIGIN

Query Match 29.1%; Score 123; DB 6; Length 290;
 Best Local Similarity 100.0%; Pred. No. 5.9e-16;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATTGATGCAAGAAATAAGAGCGCAAGATGATTCATAGCTCACTAAATTATTA 60
 Db 168 GATGATGCAAGAAATAAGAGCGCAAGATGATTCATAGCTCACTAAATTATTA 227
 QY 61 TATTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
 Db 228 TATTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
 QY 121 TGT 123
 Db 288 TGT 290

Search completed: April 15, 2003, 21:03:31
 Job time : 1213.11 secs



	Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
	8.8%	Score 37.2;	DB 10;	Length 225;			
	56.6%	Pred. No. 7.6;					
	0;		53;	Indels	0;	Gaps	0;
QY	97	AGACTCTAGCAGAAATGGCCCTGTTGTACATTATATCTCTTCCTCTAGTTGGCTGTAT	156				
Db	730	AGACTCTGCGGTTTTAGCTTGCAGAGTCAGATATTTCTCCACAGTTGGAGCATATTTT	671				
QY	157	TTCTTACTTATCTTCATTTTGGGACCTCACAGAACAATTAGCCCATTAATTCACAC	216				
Db	670	TTTCATATTTTTCCTCAAAACTGCTCCCAAAAGAACGAAAGAGATTAATTCACGAC	611				
QY	217	CT	218				
Db	610	TT	609				

```

Yr 10 GAAGAAATTAAGAGGAGAGATGATTCATATAGCTCAATAATTTATATATTTGTAT 69
Db 3121 GATTTAAAAAAGAGCTATGACCAAAATGGCTTAAAGATATTTTAAAGAGAG 3180
Yr 70 GATGATGTGAACCTCTGGAATGCTGAGACTCTAGCAGAAAATGGCTGTTTACATTT 129
Db 3181 AAAGATTAAGAAACAOTGTATATACATGAGCATATGTTGACTCTTTTCTTACACTT 3240
Yr 130 ATATCTCTTCTCTAGTTGCTGTAT 156
Db 3241 AAGCTAGATTTCTTTAAGGTATAT 3267

RESULT 6
US-09-070-927A-89
; Sequence 89, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunach
; Patrick J. Dillon
; Steven Barash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 20072 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 89:
US-09-070-927A-89

Query Match 8.7%; Score 36.6; DB 10; Length 20072;
Best Local Similarity 48.8%; Pred. No. 29;
Matches 99; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Yr 205 TAAATTCACACCTGAGAGGCTGTGTTTGAAGAGGATATGATTTATGAGAAATGATA 264
Db 17319 TAAATTTGTCCTGCTTAAAGATTTTGTAGTACCATCTGTTTAAATTTTAAAT 17378
Yr 265 TGGCAATGTGCTTACGATTTTGTATGAAAGATTTCCAGAGCTACTTCTACAGTATTTTG 324

```

```

Db 17379 TGCCACGAGGCAACATTTTCATGATATATACCATCTTGACGACAAATTTAGTCG 17438
QY 325 GTCAATATTTGGAATGCGTTTGTAGTCTTTCACCTTTTAAATTAATGTCACATACTTGTA 384
Db 17439 TTAGGTAACCGGTAAGTCTTGTTATTTAAGATATAGCCGTTGGCTTTAATTTCTG 17498
QY 385 TGAGTTCAATTAATTAATTTGACT 407
Db 17499 TAATTGCTATACACGATTTGACT 17521

RESULT 7
US-09-864-761-2816/c
; Sequence 2816, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2816
; LENGTH: 395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AI033539.17
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2

```

```

; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
US-09-864-761-2816
Query Match 8.6%; Score 36.2; DB 10; Length 395;
Best Local Similarity 57.5%; Pred. No. 6.3;
Matches 65; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 304 GCTACTTCCTACAGATATTTGGTCATATTTGGAATGCGTTTGTAGTCTTACCTTTAA 363
Db 201 GATATTTTCTAAATACCTTTTGTGAAATTTATGTGTAGACGATTTAAGTTAAAT 142
QY 364 ATTATGTCACAACTTGTATGAGTTCAATTAATTTGACTTAATGTAA 416
Db 141 ATTTATCTTAACCTTTTATGATTTAGACCAAAACATATGAGTAATGAA 89

RESULT 8
US-09-764-864-735
; Sequence 735, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 735
; LENGTH: 1841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1841)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-735
Query Match 8.5%; Score 35.8; DB 10; Length 1841;
Best Local Similarity 53.1%; Pred. No. 16;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 14 AATTAAGAGGCAAGATGATTCATAGCTCACTAAATTTATATATTTGTATGATG 73
Db 1314 AAAAAAAGAGACATGACCAAAATGGCTTAAGATTAAGTATTTTAAGGAAG 1373
QY 74 ATTTGGAACCTCTGAATGCTGAGACTCTAGACAGAAATGCGCTGTGTACATTAT 133
Db 1374 ATTAAGAACATCTTTATACATGAGCTATGCTTCTTTCTTACACTTAAGC 1433
QY 134 CTCTTCTCTTGTGCTGTAT 156
Db 1434 CTAGAAATTTCTTTAGGTATAT 1456

RESULT 9
US-10-014-799A-3
; Sequence 3, Application US/10014799A
; Publication No. US20030055219A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Cimbara, Daniel M.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-272-II

```

Query Match	8.5%	Score 35.8	DB 9	Length 513509
Best Local Similarity	50.3%	Pred. No. 1.8e+02		
Matches	88	Conservative	87	Indels 0
				Gaps 0
242	ATATGATTTTATGGAGATGATGATGATGCTTAACGATTTTGATGAAAGTTTCCC	301		

US-09-974-300-7472

Query Match 8.2%; Score 34.6; DB 10; Length 411;
Best Local Similarity 50.3%; Pred. No. 16;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 5 GATGTGAAGAAATTAAGAGCAGAGATGATTCATAGCTCACTAAATTTATATAT 64
DB 211 GATGGGATATTAAGCAGATGAGCTTGATGCTACTGTTTACCAATGCGCTTT 152
QY 65 TGTATGATGATTTGAACTCTCTGATGCTGAGACTCTAGCAGAAATGCGCTTTGTA 124
DB 151 CGTGTGGCTGATTCACCCCTAGCTCCCAAGATCTTTATATTAATGAGCTCTTCAGA 92
QY 125 CATTATATCTCTCTCTCTAGTTGGCTGATTTCTTACTTATCTTCA 173
DB 91 AACTTGATCTCTTCTTCTTACTGCGCTGACCTTGTCTTATATGCGCA 43

RESULT 13
US-09-867-701-4700/C
; Sequence 4700, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4700
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-4700

Query Match 8.2%; Score 34.6; DB 10; Length 439;
Best Local Similarity 55.4%; Pred. No. 17;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 302 AAGTACTCTACAGATTTTGGCATATTTGGAATGCGTTTACTTCACTTTT 361
DB 233 AACTATTTTGGTATTTGATTAATGAATTAATGATGACCTTTGTTTGGTTTT 174
QY 362 AAATATGCTCACTAACTTTGTATGAGTTCAATTAATATTTGACTAATGTAATGTG 421
DB 173 GAAGTATCTCTATAGCTTGTGATGACTGTATGAGAAACTAGCTAATGTAATAG 114
QY 422 A 422
DB 113 A 113

RESULT 14
US-09-920-455-206/C
; Sequence 206, Application US/09920455
; Patent No. US20020168647A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.540
; CURRENT APPLICATION NUMBER: US/09/920,455
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 275
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 455
; OTHER INFORMATION: n = A,T,C or G
US-09-920-455-206

Query Match 8.2%; Score 34.6; DB 9; Length 461;
Best Local Similarity 49.2%; Pred. No. 17;
Matches 91; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 12 AGAATTAAGAGGAGAGATGATTCATAGCTCACTAAATTTATATTTATGTA 71
DB 451 AAATGATGATTTGAGATATTAATGAGAAATTTCTCAAAATGATGATTTAGGAAA 392
QY 72 TGATTTGAACCTCTGAATGCCGAGACTCTAGCAGAAATGCGCTTTGTATTTAT 131
DB 391 TGATTTGCTTCTCTTAATAGTTGAGAGGCTGAAATTAACCTTTTGTGATTTCT 332
QY 132 ATCTCTCTCTAGTGGCTGATTTCTTACTTATCTTCACTTTTGGCAGCTCAGAGA 191
DB 331 TTTGAAATGTTTGGTCAATTACAACTTTTAACCTTATCTTCTCTTACGCCCTTA 272
QY 192 ACAA 196
DB 271 ACAGA 267

RESULT 15
US-10-121-746-26/C
; Sequence 26, Application US/10121746
; Publication No. US20030036648A1
; GENERAL INFORMATION:
; APPLICANT: Miller, Andrew P.
; APPLICANT: Curran, Mark Edward
; APPLICANT: Hu, Ping
; APPLICANT: Rutler, Marc
; APPLICANT: Wang, Jian-Wang
; TITLE OF INVENTION: NO. US20030036648A1 Human Potassium Channels
; FILE REFERENCE: SEQ-15P
; CURRENT APPLICATION NUMBER: US/10/121,746
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: US/09/336,643A
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/076,687
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/116,448
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/US99/03826
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 3204
; TYPE: DNA
; ORGANISM: H. sapiens
FEATURE:
; NAME/KEY: CDS
; LOCATION: (182)...(1349)
; OTHER INFORMATION: K+Hov42
US-10-121-746-26

Query Match 8.2%; Score 34.6; DB 9; Length 3204;
Best Local Similarity 55.4%; Pred. No. 41;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 302 AAGTACTCTACAGATTTTGGCATATTTGGAATGCGTTTACTTCACTTTT 361
DB 2277 AACTATTTTGGTATTTGATTAATGAATTAATGATGACCTTTGTTTGGTTTT 2218
QY 362 AAATATGCTCACTAACTTTGTATGAGTTCAATTAATATTTGACTAATGTAATGTG 421
DB 2217 GAAGTATCTCTATAGCTTGTGATGACTGTATGAGAAACTAGCTAATGTAATAG 2158

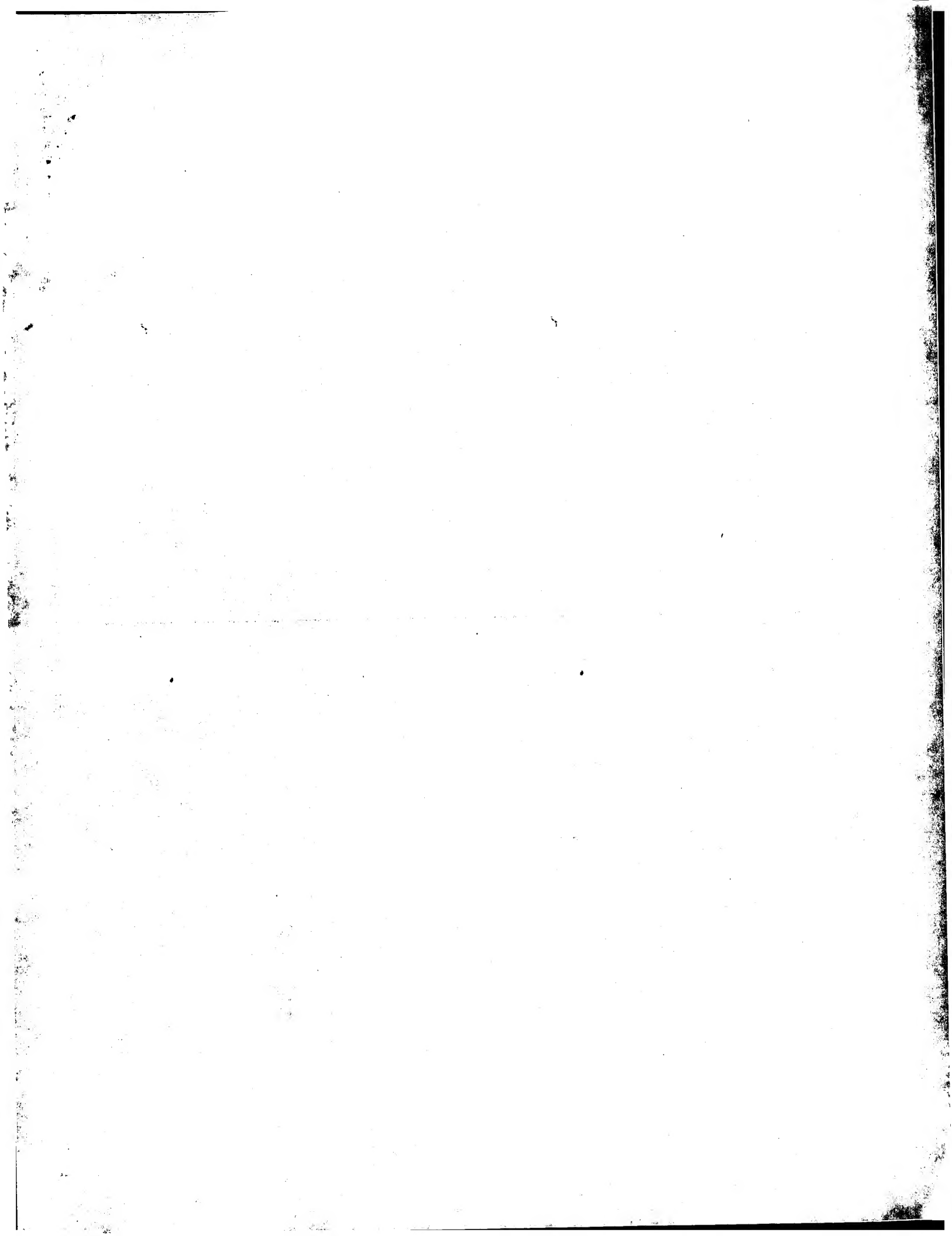
Wed Apr 16, 05:42:58 2003

us-09-647-019-10.rnpb

Page 7

Qr 422 A 422
Db 2157 A 2157

Search completed: April 16, 2003, 01:19:58
Job time : 187.167 secs



C	2	407	96.4	909	4	US-09-484-9270B-111	Sequence 111, App
C	1	39.4	9.3	14672	4	US-08-961-527-111	Sequence 111, App
C	3	38.8	9.2	1755	4	US-09-134-001C-1047	Sequence 111, App
C	4	37.2	8.8	25002	4	US-08-961-527-48	Sequence 1047, App
C	5	35.4	8.4	3182	4	US-08-971-395-1	Sequence 48, App
C	6	35.4	8.4	3183	1	US-08-413-135-1	Sequence 1, App1
C	7	34.6	8.2	3204	4	US-09-336-643A-66	Sequence 1, App1
C	8	33.4	7.9	1934	4	US-09-071-035-45	Sequence 26, App
C	9	33.4	7.9	3963	4	US-09-071-035-45	Sequence 45, App
C	10	33.4	7.9	3963	4	US-09-071-035-45	Sequence 45, App
C	11	33.4	7.9	3963	4	US-09-071-035-45	Sequence 45, App
C	12	33.2	7.9	860	4	US-08-557-146-1	Sequence 287, App
C	13	33.2	7.9	969	4	US-09-502-600-30	Sequence 287, App
C	14	33.2	7.9	986	2	US-08-557-146-1	Sequence 10, App
C	15	33.2	7.9	986	2	US-09-154-344-1	Sequence 1, App1
C	16	33.2	7.9	1089	3	US-08-930-188-1	Sequence 1, App1
C	17	33.2	7.9	1089	3	US-08-930-188-1	Sequence 1, App1
C	18	33.2	7.9	1089	3	PCT-US96-04294-3	Sequence 3, App1
C	19	33.2	7.9	1089	5	PCT-US96-04294-3	Sequence 1, App1
C	20	32.8	7.8	1121	6	US-08-687-080-105	Sequence 3, App1
C	21	32.8	7.8	4304	4	5265558-7	Sequence 105, App
C	22	32.8	7.8	4304	5	US-08-368-776A-1	Patent No. 5265558
C	23	32.6	7.7	9636	1	PCT-US96-00419-1	Sequence 1, App1
C	24	32.6	7.7	9636	1	US-08-323-170B-1	Sequence 1, App1
C	25	32.4	7.7	6946	4	US-08-954-441-1	Sequence 1, App1
C	26	32.2	7.6	4456	4	US-09-316-080-1	Sequence 1, App1
C	27	32.2	7.6	5173	1	US-09-095-443-1	Sequence 1, App1
C	28	32.2	7.6	5173	1	US-08-242-677-1	Sequence 1, App1

28	32	7.6	2674	2	US-08-976-124-2	Sequence 2,	Appl1
29	31.8	7.5	1359	1	US-07-618-1124-1	Sequence 1,	Appl1
30	31.8	7.5	1359	1	US-08-110-7864-7	Sequence 7,	Appl1
31	31.8	7.5	1359	1	US-08-200-4228-1	Sequence 1,	Appl1
32	31.8	7.5	1512	2	US-08-909-9655-8	Sequence 8,	Appl1
33	31.8	7.5	1852	1	US-08-638-8114-6	Sequence 6,	Appl1
34	31.8	7.5	1852	1	US-08-332-663-14	Sequence 14,	Appl1
35	31.4	7.4	1844	4	US-09-078-1994-6	Sequence 6,	Appl1
36	31.2	7.4	2605	2	US-08-680-4995-4	Sequence 4,	Appl1
37	31	7.3	591	4	US-09-385-982-72	Sequence 72,	Appl1
38	31	7.3	2979	4	US-09-437-277-4	Sequence 4,	Appl1
39	31	7.3	12124	1	US-08-181-2713-36	Sequence 36,	Appl1
40	31	7.3	12124	1	US-08-449-315-36	Sequence 36,	Appl1
41	31	7.3	12124	1	US-08-444-803-36	Sequence 36,	Appl1
42	31	7.3	12124	1	US-08-449-043-36	Sequence 36,	Appl1
43	31	7.3	12124	1	US-08-456-2659-36	Sequence 36,	Appl1
44	31	7.3	12124	1	US-08-455-415-36	Sequence 36,	Appl1
45	31	7.3	12124	1	US-08-455-244-36	Sequence 36,	Appl1

ALIGNMENTS

RESULT 1

```

US-09-484-970B-111
// Sequence 111, Application US/09484970B
// Patent No. 6426186
// GENERAL INFORMATION:
// APPLICANT: Jones, Karen A.
// APPLICANT: Volkmuth, Wayne
// APPLICANT: Walker, Michael G.
// TITLE OF INVENTION: BONE REMODELING GENES
// FILE REFERENCE: PB-0014 US
// CURRENT APPLICATION NUMBER: US/09/484,970B
// CURRENT FILING DATE: 2000-01-18
// NUMBER OF SEQ ID NOS: 172
// SOFTWARE: PERL Program
// SEQ ID NO 111
// LENGTH: 909
// TYPE: DNA
// ORGANISM: Homo sapiens
// FEATURE:
// NAME/KEY: misc feature
// OTHER INFORMATION: Incyte ID No. 6426186 021656-2CB3
US-09-484-970B-111

```

Query Match	96.4%;	Score 407;	DB 4;	Length 909
Best Local Similarity	100.0%;	Pred No 1	Eq 100	

Matches	407;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0
---------	------	--------------	----	------------	----	--------	----	------	---

OY	1	GATTGATGTGTAGAAAATAAAGAGCAGAAATGATTCATTAAGCTCATTAATTTTATA	60
Db	503	GATTGATGTGTAGAAAATAAAGAGCAGAAATGATTCATTAAGCTCATTAATTTTATA	56
OY	61	TATTTGATGATGATTTGTGTAACTCTGAAATGCTGAGACTTAGCAGAAATGGCCTGTT	12
Db	563	TATTTGATGATGATTTGTGTAACTCTGAAATGCTGAGACTTAGCAGAAATGGCCTGTT	62
OY	121	TGTACATTTATATCTCTCTCTCTAGTGGCTGTAAATTTCTACTTTATCTTCATTTTGG	18
Db	623	TGTACATTTATATCTCTCTCTCTAGTGGCTGTAAATTTCTACTTTATCTTCATTTTGG	68
OY	181	CACCTCACGAAACAATTAGCCCATTAATTCACAACCTGAGGGGTGTGTTTGAGAGAG	24
Db	683	CACCTCACGAAACAATTAGCCCATTAATTCACAACCTGAGGGGTGTGTTTGAGAGAG	74
OY	241	GATATGATTTTATGAGGAATGATATGCAATGTGCGCTAACGATTTTGATGAAGAATTTCC	30
Db	743	GATATGATTTTATGAGGAATGATATGCAATGTGCGCTAACGATTTTGATGAAGAATTTCC	80
OY	301	CAAGCTACTTCTCTACAGTATTTTGTCAATATTTGGAAATGCGTTTAACTCTTCAACCTT	36
Db	803	CAAGCTACTTCTCTACAGTATTTTGTCAATATTTGGAAATGCGTTTAACTCTTCAACCTT	86

INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 25002 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-48

Query Match 8.4%; Score 37.2; DB 4; Length 25002;
Best Local Similarity 56.6%; Pred. No. 0.78; Indels 53; Gaps 0;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 97 AGACTCAGCAAGAAATGCGCTGTTGTACATTATCTCTCTAGTTGCTGTAT 156
DB 8192 AGACTCTGCGGTTTACGTTCAGAGCCAGTAATTTCTCCACAGCTGGAGTATTT 8251
QY 157 TTCTTACTTATCTTCTTCTTGGACCTCAGCAAGAAATTTGCCATTAATCAAC 216
DB 8252 TTCTCATTTTCTCTCAAAAGCTGCTCCAAAGAAAGAGATTAATTCAGCAC 8311
QY 217 CT 218
DB 8312 TT 8313

RESULT 5
US-08-971-395-1
Sequence 1, Application US/08971395

GENERAL INFORMATION:
APPLICANT: Amasino, Richard M
APPLICANT: Gan, Susheng
TITLE OF INVENTION: Transgenic Plants with Altered
NUMBER OF SEQUENCES: 5
TITLE OF INVENTION: Senescence Characteristics
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pluckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971.395
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.94908
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3182 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-971-395-1

Query Match 8.4%; Score 35.4; DB 4; Length 3182;
Best Local Similarity 48.3%; Pred. No. 1.3; Indels 106; Gaps 0;
Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 103 TAGCAGAAATGCGCTGTTGTACATTATCTCTCTAGTTGCTGTATTTCTTA 162

DB 197 TAGCTAATATAGCAATGATTTAAATTTGTTTGAACACCTTTTCTCTCTGTCG 256
QY 163 CTTTATCTTCAATTTTGGCACTCAGCAAGAAATTTGCCATTAATTCACACCTGGAG 222
DB 257 TTTTCTTACATTTAGAAAGAACCCATTAACATGTCGTTCAATTAATTAACAAATATT 316
QY 223 GGTGTGTTTGAAGAGGATATGATTTTATGAGAAATGATATGCAATGCTTACGA 282
DB 317 TCCAAGTTTATATACGAAACCTTGTTTATGAAACAGTTGAATGATGATATGA 376
QY 283 TTTTGAGCAAAAGTTTCCCAAGCTA 307
DB 377 ATTAGTATGATCAATTAATCAATATA 401

RESULT 6
US-08-413-135-1
Sequence 1, Application US/08413135

GENERAL INFORMATION:
APPLICANT: Amasino, Richard M
APPLICANT: Gan, Susheng
TITLE OF INVENTION: Transgenic Plants with Altered
NUMBER OF SEQUENCES: 3
TITLE OF INVENTION: Senescence Characteristics
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pluckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413.135
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.92808
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-9166
TELEFAX: 608-251-5000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3183 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "SMG12-1 Promoter DNA"
US-08-413-135-1

Query Match 8.4%; Score 35.4; DB 1; Length 3183;
Best Local Similarity 48.3%; Pred. No. 1.3; Indels 106; Gaps 0;
Matches 99; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 103 TAGCAGAAATGCGCTGTTGTACATTATCTCTCTAGTTGCTGTATTTCTTA 162
DB 197 TAGCTAATATAGCAATGATTTAAATTTGTTTGAACACCTTTTCTCTCTGTCG 256
QY 163 CTTTATCTTCAATTTTGGCACTCAGCAAGAAATTTGCCATTAATTCACACCTGGAG 222
DB 257 TTTTCTTACATTTAGAAAGAACCCATTAACATGTCGTTCAATTAATTAACAAATATT 316
QY 223 GGTGTGTTTGAAGAGGATATGATTTTATGAGAAATGATATGCAATGCTTACGA 282

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brooks
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: FB369P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 449:
SEQUENCE CHARACTERISTICS:
LENGTH: 3963 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-449

Query Match 7.9%; Score 33.4; DB 4; Length 3963;
Best Local Similarity 51.4%; Pred. No. 4.8;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 51 AAATTTATATATTTGTATGATGATGTAACCTCTGAATGCTGAGACTTACAGAA 110
Db 2582 AAAGCTTGNATATCCGCGACTCTCTTATTTCTTATACAGCTGCTTCTGGA 2523
Qy 111 ATGGCTGTTGTACATTATATCTCTCTTCTAGTTGGCTGATTTCTTACTTATCT 170
Db 2522 ATGGGCTTTTAAATGTCACCTTCTTACGTTTTCGGGTGTGTTTGGTGTCTTCT 2463
Qy 171 TCATTTTGGCACCCTCACAGACAATT 198
Db 2462 GGGGTCTGCACACTACATGCGCAACT 2435

RESULT 10

US-09-071-035-453/c
Sequence 453, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:

APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION/DOCKET NUMBER: 36,373
TELECOMMUNICATION INFORMATION: PB369P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 453:
SEQUENCE CHARACTERISTICS:
LENGTH: 3963 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-453

Query Match 7.9%; Score 33.4; DB 4; Length 3963;
Best Local Similarity 51.4%; Pred. No. 4.8;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 51 AAATTTATATATTTGTATGATGATGTAACCTCTGAATGCTGAGACTTACAGAA 110
Db 2582 AAAGCTTGNATATCCGCGACTCTCTTATTTCTTATACAGCTGCTTCTGGA 2523
Qy 111 ATGGCTGTTGTACATTATATCTCTCTTCTAGTTGGCTGATTTCTTACTTATCT 170
Db 2522 ATGGGCTTTTAAATGTCACCTTCTTACGTTTTCGGGTGTGTTTGGTGTCTTCT 2463
Qy 171 TCATTTTGGCACCCTCACAGACAATT 198
Db 2462 GGGGTCTGCACACTACATGCGCAACT 2435

RESULT 11

US-08-998-416-287
Sequence 287, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:

APPLICANT: Phillippen, Peter
APPLICANT: Polmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Reibschung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998.416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION/DOCKET NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 287:
SEQUENCE CHARACTERISTICS:
LENGTH: 860 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1240UP
US-08-998-416-287

Query Match 7.9%; Score 33.2; DB 4; Length 860;
Best Local Similarity 48.4%; Pred. No. 3.5;
Matches 89; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 230 TTTTGAGAGGATATGATTTTATGAGATGATGCAATGTGCTTAAGATTTTAT 289
Db 534 TTTTAAATGTTTATGTTTAAATGATTAATACATTAATAATATAAATTAAGATGCC 593

QY 290 GAAAGTTTCCCAAGTACTCTCCACAGTATTTGGTCAATATTGGATGCGTTTAGT 349
DB 594 ACAATATATCCCATTTTCCCTTATGATCAATTAATGTTTCCATTTATTTTACTA 653
QY 350 TCTTCACTTTTAATTAATGTCATAAAGTTTGTATGAGTTCAATAATATTGACTAA 409
DB 654 TTTTATCCTTTATCTATATGATGTTTTTACCTAAGATTTAANAATATATATCTCTAATAT 713
QY 410 ATGT 413
DB 714 ATAT 717

RESULT 12
US-09-502-600-30
Sequence 30, Application US/09502600A
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
FILE REFERENCE: D6223CIP-C
CURRENT FILING DATE: 2000-02-11
CURRENT APPLICATION NUMBER: US/09/502,600A
PRIOR FILING DATE: 03-14-1998
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 30
LENGTH: 969
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: full length cDNA of SCCE
US-09-502-600-30

Query Match 7.9%; Score 33.2; DB 4; Length 969;
Best Local Similarity 57.8%; Pred. No. 3.6; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 43

QY 317 GTATTGTGTCATATTGGAATGCGTTTATGTTCTTCACCTTTTAATTATGTCATA 376
DB 7 GTTTTGCTTTCTTTATTTGTTTGGTTTAGGTCTTACCAATTGATGGTTATCA 66

QY 377 ACTTTGATAGTTCATAATATATTGACTAATGTAAT 418
DB 67 CAGGCATGAGTTTAATATATCTTTGAGGAAGGTAAGT 108

RESULT 13
US-08-557-146-1/c
Sequence 1, Application US/08557146
GENERAL INFORMATION:
PATENT NO. 5834290
APPLICANT: Egelund, Torbjorn
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,146

FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 986 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 25..786
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 25..90
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 112..783
US-08-557-146-1

Query Match 7.9%; Score 33.2; DB 2; Length 986;
Best Local Similarity 57.8%; Pred. No. 3.6; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 43

QY 317 GTATTGTGTCATATTGGAATGCGTTTATGTTCTTCACCTTTTAATTATGTCATA 376
DB 972 GTTTTGCTTTCTTTATTTGTTTGGTTTAGGTCTTACCAATTGATGGTTATCA 913

QY 377 ACTTTGATAGTTCATAATATATTGACTAATGTAAT 418
DB 912 CAGGCATGAGTTTAATATATCTTTGAGGAAGGTAAGT 871

RESULT 14
US-09-154-344-1/c
Sequence 1, Application US/09154344
GENERAL INFORMATION:
PATENT NO. 5981256
APPLICANT: Egelund, Torbjorn
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,344
FILING DATE: 16-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,146
FILING DATE: 14-DEC-1995

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 986 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 25..786
NAME/KEY: 819_peptide
LOCATION: 25..90
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 112..783
US-09-154-344-1

Query Match 7.9%; Score 33.2; DB 2; Length 986;
Best Local Similarity 57.8%; Pred. No. 3.6;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 317 GATTTTGGTCAATATTGGAAATGCGTTTACCTTCACCTTTAAATTAATGTCACCTAA 376
Db 972 GTTTTGTGTTCTTATTGTTGTTGTTAGTCTTACCAATTTGATTTGTTATCA 913

Qy 377 ACTTTGATGAGTTCAAATTAATATTGACTAAATGTAAT 418
Db 912 CAGGGCATGAGTTTAAATATATCTTTGAGGAAGTAAGT 871

RESULT 15
US-08-930-188-1
Sequence 1, Application US/08930188
Patent No. 6093397
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
TITLE OF INVENTION: RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,188
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Blalock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1089 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-930-188-1

Query Match 7.9%; Score 33.2; DB 3; Length 1089;
Best Local Similarity 57.8%; Pred. No. 3.8;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 317 GATTTTGGTCAATATTGGAAATGCGTTTACCTTCACCTTTAAATTAATGTCACCTAA 376
Db 26 GTTTTGTGTTCTTATTGTTGTTGTTAGTCTTACCAATTTGATTTGTTATCA 85

Qy 377 ACTTTGATGAGTTCAAATTAATATTGACTAAATGTAAT 418
Db 86 CAGGGCATGAGTTTAAATATATCTTTGAGGAAGTAAGT 127

Search completed: April 15, 2003, 22:53:50
Job time : 46.9694 secs

